

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»
(Самарский университет)

Н.А. САЗОННИКОВА

УПРАВЛЕНИЕ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ

Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» в качестве учебного пособия для студентов, обучающихся по основной образовательной программе высшего образования по направлению подготовки 28.03.02 Наноинженерия

© Самарский университет, 2017

ISBN 978-5-7883-1197-5

САМАРА
Издательство Самарского университета
2017

УДК 621.002(075)

ББК 34.5я7

С 148

Рецензенты: д-р тех. наук, проф. А.Л. Кривченко,
д-р тех. наук, проф. Н.Д. Проничев

С 148 **Сазонникова, Надежда Александровна**

Управление технологическими процессами: учеб. пособие / *Н.А. Сазонникова*. – Электрон. текст. и граф. дан. (4,25 Мб). – Самара: Издательство Самарского университета, 2017. – 6 опт. компакт-дисков (CD-ROM). – Систем. требования: ПК Pentium, Adobe Acrobat Reader. – Загл. с титул. экрана.

ISBN 978-5-7883-1197-5

В данном пособии рассмотрены методы математического моделирования и методика проведения экспериментальных исследований применительно к исследованию технологических процессов, теоретические основы однофакторного и многофакторного корреляционного и регрессионного анализа. Приведены порядок выбора параметров оптимизации и факторов технологических процессов, методы составления матрицы планирования эксперимента, проведения статистического анализа уравнения регрессии. Описаны планы первого и второго порядков, методы оптимизации технологических процессов.

Предназначено для студентов, обучающихся по направлению подготовки 28.03.02 Наноинженерия по профилю «Нанотехнология и наноматериалы».

Подготовлено на кафедре автоматических систем энергетических установок.

УДК 621.002(075)

ББК 34.5я7

© Самарский университет, 2017

Редактор Ю.Н. Литвинова
Компьютерная верстка Ю.Н. Литвинова

Подписано для тиражирования 27.11.2017.
Объем издания 4,25 Мб.
Количество носителей 6 экз.

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ имени академика С. П. КОРОЛЕВА»
(Самарский университет)
443086, Самара, Московское шоссе, 34.

Изд-во Самарского университета.
443086 Самара, Московское шоссе, 34.

ОГЛАВЛЕНИЕ

Введение	5
1. Место и роль математического моделирования в задачах исследования, проектирования и оптимизации технологических систем	7
1.1. Использование моделирования при исследовании и проектировании АСУ	9
1.2. Классификация процессов как объектов моделирования	10
1.3. Постановка задачи моделирования в общем виде	11
1.4. Общие сведения о математическом моделировании	13
1.5. Структурно-параметрическое описание и назначение параметров объекта	19
1.6. Дискретные и непрерывные модели	22
2. Математический подход к анализу и оптимизации технологических процессов в машиностроении	24
2.1. Основные понятия и определения	24
2.2. Определение и выбор параметров оптимизации и факторов технологических процессов	28
Контрольные вопросы	36
3. Рациональное планирование	38
Контрольные вопросы	44
4. Корреляционно-регрессионный однофакторный анализ	45
Контрольные вопросы	57
5. Проверка статистических гипотез	58
5.1. Понятие статистической гипотезы	58
5.2. Теоретическая схема проверки гипотез	59
5.3. Проверка типовых статистических гипотез	64
Контрольные вопросы	92
6. Выбор модели многофакторного регрессионного анализа	93
6.1. Предпланирование эксперимента	93
6.2. Шаговая процедура предпланирования	97
6.3. Интерполяционная модель	99
Контрольные вопросы	102
7. Полный факторный эксперимент. Построение регрессионной модели	103
Контрольные вопросы	115
8. Принятие решений после построения модели технологического процесса	117
Контрольные вопросы	122
9. Планы дробного факторного эксперимента	123
Контрольные вопросы	128
10. Оптимизация технологических процессов	129
10.1. Понятие о задаче условной оптимизации. Классификация задач оптимизации	129
10.2. Крутое восхождение по поверхности отклика. Понятие градиента	136
10.3. Способы градиентной оптимизации	137
10.4. Особенности применения градиентной оптимизации совместно с методами планирования экспериментов	140
Контрольные вопросы	143

11. Критерии оптимальности и типы планов	144
Контрольные вопросы.....	147
12. Планы второго порядка	148
12.1. Ортогональный центрально-композиционный план второго порядка	149
12.2. Пример плана ОЦКП для $n = 2$	154
12.3. Ротатабельные планы	155
12.5. Пример ротатабельного ортогонального центрально-композиционного плана для $n = 2$	159
12.6. Планы второго порядка с единичной областью планирования	160
12.7. Ротатабельный план на основе правильного многоугольника при $n=2$	162
Контрольные вопросы.....	165
13. Корреляционный анализ	166
13.1. Парная корреляция	166
13.2. Многомерный корреляционный анализ	173
13.3. Ранговая корреляция Спирмена.....	176
Контрольные вопросы.....	179
СПИСОК ЛИТЕРАТУРЫ.....	180
ПРИЛОЖЕНИЯ	182

ВВЕДЕНИЕ

Интенсивное развитие авиационной, космической и других отраслей техники существенно повышает требования к технологии как науке о различных сторонах производственных процессов.

Специальные исследования показывают, что технологию целесообразно рассматривать как стохастическую систему, в которой действуют детерминированные, т.е. точно определенные закономерности. Подход к исследованиям технологии как стохастической системы приводит к необходимости изучать статистические закономерности взаимосвязи технологических и эксплуатационных характеристик изделий.

В результате технического прогресса технологические рекомендации, основанные только на качественной оценке или описывающие зависимость свойств от единичного технологического фактора, в настоящее время не дают оптимальных практических результатов. Решение проблемы анализа и оптимизации технологии в таких условиях целесообразно вести, опираясь на строгие математические методы, которые, не заменяя собой традиционных в технологии физических, химических и других методов познания, дополняют и развивают их, обеспечивая качественно новый уровень технологических исследований.

Цель данного пособия – в достаточно доступной форме познакомить студентов с основными идеями и методами управления технологическими процессами на основе математического моделирования и дать методику проведения экспериментальных исследований применительно к исследованию технологических процессов производств, а также к новым технологическим процессам изготовления деталей и их сборки.

1. МЕСТО И РОЛЬ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ В ЗАДАЧАХ ИССЛЕДОВАНИЯ, ПРОЕКТИРОВАНИЯ И ОПТИМИЗАЦИИ ТЕХНОЛОГИЧЕСКИХ СИСТЕМ

Для изучения и оптимизации существующих и проектирования вновь Разрабатываемых технологических систем их необходимо представить в виде, удобном для исследования. С этой целью техническую систему упрощают. Такую систему (или ее часть), называемую системой автоматического управления (СУ), можно представить состоящей из двух частей: объекта управления (ОУ) и устройства управления (УУ) (рис. 1.1, а). На рис. 1.1, б представлен пример системы управления процессом плавки.

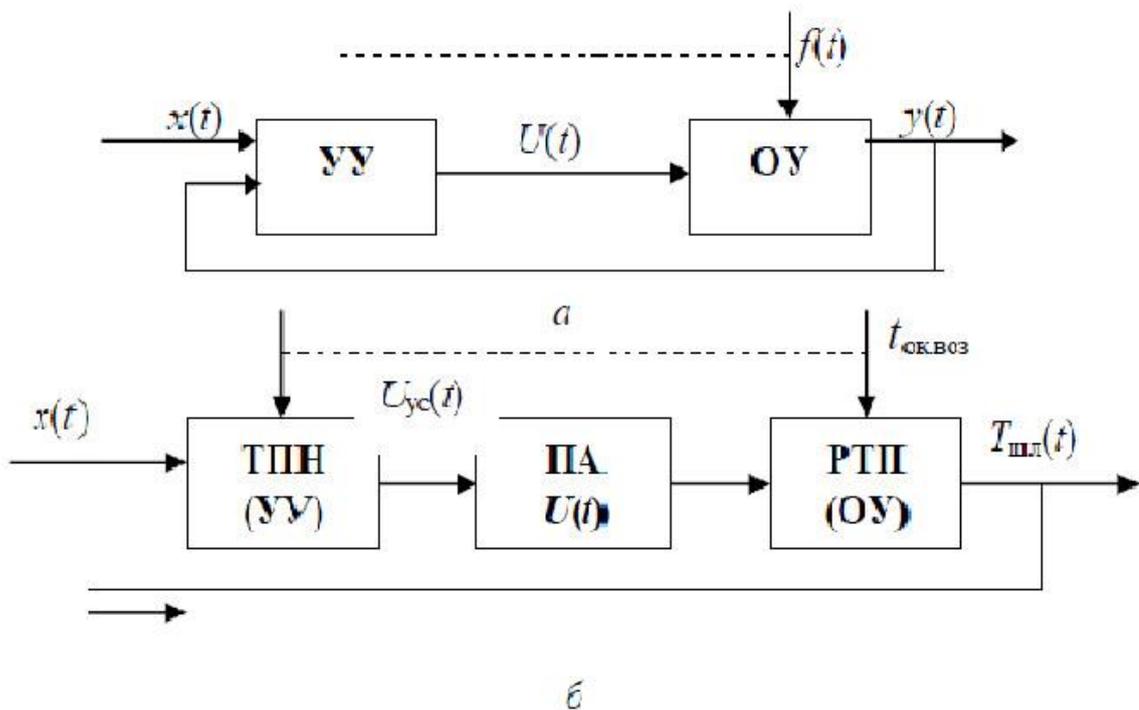


Рис. 1.1. Структурная схема системы автоматического моделирования:

а – в общем виде; б – система автоматического управления электрическим режимом технологического РТП

Здесь ТПН – это тиристорный преобразователь напряжения; ПА – перемещение электрода; $x(t)$ – входное воздействие; $y(t)$ – выход системы, ОУ; $U_{yc}(t)$ – управляющее воздействие; $f(t)$ – случайные возмущения; РТП – рудно-термическая печь; $T_{шл}(t)$ – температура шлака; $t_{ок.воз}$ – случайное возмущение температуры окружающего воздуха.

Под ОУ применительно к задачам инженерии подразумевается техническое устройство, процессом $y(t)$ на выходе которого надлежит управлять, УУ обобщает все входящие в контур СУ элементы, используемые с целью организации процесса управления. На вход системы управления подается воздействие $x(t)$, определяющее желаемый характер управляемого процесса $Y(t)$, УУ на основании информации о процессах $x(t)$ и $y(t)$, а в ряде случаев и на основании данных о возмущениях $f(t)$, рассчитывает управляющее воздействие $U(t)$ на объект с целью приведения процесса $y(t)$ в соответствие с сигналом $x(t)$.

Для решения большинства задач анализа и синтеза СУ необходимо иметь математическую модель ОУ. Построение математической модели заключается в установлении ряда соотношений, позволяющих при каждом входном воздействии и начальных состояниях найти сигнал на выходе ОУ.

Обычно модель получают как математическую формулировку физических законов, которым подчинена работа ОУ. В общем случае ОУ является многомерным и имеет l управляемых процессов – $y_1(t), y_2(t), \dots, y_l(t)$, m управляющих воздействий – $u_1(t), u_2(t), \dots, u_m(t)$, k внешних возмущений – $f_1(t), f_2(t), \dots, f_k(t)$ (рис. 1.2).

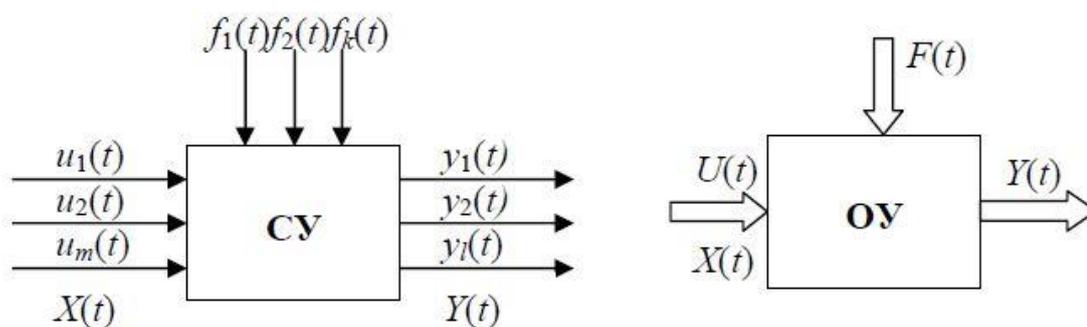


Рис. 1.2. Структурная схема управления объектом моделирования

Математическая запись физических законов, определяющих свойства непрерывного объекта, в большинстве случаев может быть представлена как система нелинейных дифференциальных уравнений, связывающих входные и выходные процессы и их производные:

$$\Phi_i \begin{bmatrix} y_1(t), y_1^{(1)}(t), \dots, y_1^{(n)}(t), \dots, y_i^{(n)}(t); \\ u_1, u_1^{(n)}, \dots, u_1^{(n)}, u_2(t), \dots, u_2^{(n)}, u_m(t), \dots, u_m^{(n)}; \\ f_1, f_1^{(1)}, \dots, f_1^{(n)}, f_2^{(1)}, \dots, f_2^{(n)}, f_m(t), \dots, f_m^{(n)} \end{bmatrix} = 0$$

или

$$\Phi_i[y(t), u(t), f(t)] = 0, \quad i = 1, 2, \dots, l. \quad (1.1)$$

При $l = 1$ объект называют одномерным.

Если функции Φ_i являются линейными относительно управляемых и управляющих процессов и их производных, то объект называют линейным по управлению, аналогично линейным по возмущению.

Математическая модель (1.1) в современной теории оптимальных и Адаптивных систем получила ограниченное распространение. Гораздо чаще дифференциальные уравнения (1.1) n -го порядка представляют в виде системы из nl дифференциальных уравнений первого порядка, разрешенных относительно производных.

1.1. Использование моделирования при исследовании и проектировании АСУ

Одна из проблем современной науки и техники – разработка и внедрение в практику проектирования новейших методов исследования характеристик автоматизированных систем управления (АСУ) различных уровней. При проектировании обеспечивающих и функциональных подсистем таких АСУ возникают многочисленные задачи, требующие оценки количественных и качественных закономерностей процессов функционирования систем, проведения их структурного, алгоритмического и параметрического синтеза.

Ограниченность возможностей экспериментального исследования больших систем делает актуальной задачу разработки методики их моделирования, которая бы позволила в соответствующей форме представить процессы функционирования систем, описать протекание этих процессов с помощью математической модели, получить результаты экспериментов с моделями по оценке характеристик исследуемого объекта.

Независимо от того, какие подсистемы составляют АСУ, при *разработке* каждой из них необходимо выполнить внешнее проектирование (макропроектирование) и внутреннее проектирование (микропроектирование).

На стадии макропроектирования должна быть создана обобщенная модель процесса функционирования АСУ, позволяющая разработчику сделать выводы об эффективности различных стратегий управления объектом при его взаимодействии с внешней средой. Например, управление энергетическим режимом электропечи может быть осуществлено с помощью изменения подводимого к электродам напряжения, заглубления электродов или регулированием количества и химического состава загружаемой в электропечь шихты.

На стадии микропроектирования разрабатывают модели с целью создания эффективных обеспечивающих подсистем АСУ. Выбор метода моделирования и необходимая детализация моделей существенно зависят от этапа разработки АСУ.

На этапах *разработки технического и рабочего проектов АСУ* модели отдельных подсистем детализируются, и моделирование служит для решения конкретных задач проектирования. Целевое назначение моделирования на *этапе внедрения и эксплуатации АСУ* – это проигрывание возможных ситуаций для принятия обоснованных и перспективных решений по управлению объектом. Моделирование также широко применяют при обучении и тренировке технологического персонала АСУ. В этом случае моделирование носит характер деловых игр.

1.2. Классификация процессов как объектов моделирования

Процессы, для управления которыми создаются АСУ, можно разделить на три группы:

- ◆ *непрерывные*, которые характеризуются непрерывным режимом работы (электролиз алюминия, электроплавка никелево-медного агломерата в РТП, спекание нефелинового концентрата);

- ◆ *полунепрерывные* (непрерывно-дискретные), которые характеризуются полунепрерывным режимом работы. Например, при плавке оловянных

концентратов в РТП происходит непрерывная загрузка шихты в течение определенного промежутка времени и затем выпуск продуктов плавки;

♦ *периодические*, которые характеризуются дискретным режимом работы, например, периодической загрузкой материалов в печь, проведением технологического процесса и выгрузкой полученного продукта (выращивание монокристаллов кремния).

В соответствии с приведенной классификацией процессов выделяются АСУ непрерывными, полунепрерывными и периодическими процессами. Наибольшую сложность представляет автоматизация периодических процессов. Для реализации АСУ необходимо промоделировать описание данного технологического процесса.

1.3. Постановка задачи моделирования в общем виде

С развитием системных исследований и расширением экспериментальных методов изучения реальных объектов большое значение приобретают математические методы анализа и синтеза. Подобие и моделирование позволяют по-новому описать реальный процесс и упростить экспериментальное его изучение.

Моделирование базируется на некоторой аналогии реального и мысленного эксперимента (выявление влияния изменения химического состава шихты на технико-экономические показатели (ТЭП) процесса плавки). Для объяснения реальных процессов выдвигаются гипотезы, с целью их подтверждения ставится эксперимент, т.е. некая процедура организации и наблюдения явлений, которую осуществляют в условиях, близких к реальным или имитирующих их. Токораспределение при электролизе расплавов солей можно оценить на моделях, использующих электролиз водных растворов электролитов.

В основе любого вида моделирования лежит модель, имеющая соответствие, базирующееся на общем качестве, которое характеризует реальный объект (например, описание с помощью дифференциальных уравнений процессов массопереноса).

В основе моделирования лежат информационные процессы, поскольку самосоздание модели базируется на информации о реальном объекте. В процессе

реализации модели одновременно получается информация об ОУ, которая сравнивается с моделью, и на основе данных сравнения вырабатывается управляющее воздействие на процесс. Поэтому можно сказать, что реализация модели осуществляется одновременно с процессом.

При постановке задачи моделирования можно выделить следующие характерные признаки математической модели: цель функционирования, сложность, целостность, неопределенность, поведенческая стратегия, адаптивность, организационная структура, управляемость, возможность развития.

Цель функционирования определяет степень целенаправленного поведения модели. По этому признаку модели могут быть разделены на одно- и многоцелевые.

Сложность модели можно оценить по общему числу элементов в системе и связей между ними.

Целостность указывает на то, что создаваемая модель является одной общей системой, включает в себя большое количество составных частей, находящихся в сложной взаимосвязи друг с другом.

Неопределенность, которая проявляется в системе, оценивается энтропией. Используя эту характеристику, в ряде случаев можно определить количество управляющей информации для достижения заданного состояния системы.

Поведенческая стратегия позволяет оценить эффективность достижения системой поставленной цели. В зависимости от наличия случайных возмущений можно различать детерминированные и стохастические системы, по своему поведению – непрерывные и дискретные.

Адаптивность – приспособляемость к различным внешним возмущающим факторам в широком диапазоне изменения воздействий внешней среды. При этом система управления должна компенсировать изменение случайных факторов. Например, АСУ теплового режима электропечи должна стабилизировать температуру расплава как при изменении влажности шихты в рабочем диапазоне, так и при изменении содержания олова в концентрате, загружаемом в электропечь.

Организационная структура системы моделирования во многом зависит от сложности модели и степени совершенства средств моделирования.

Управляемость модели дает возможность обеспечивать управление процессом в различных условиях, имитирующих реальные. К этому можно отнести управление технологическим процессом как в нормальном, так и в предаварийном состоянии.

Возможность развития модели позволяет создавать мощные системы моделирования для исследования многих сторон функционирования реального объекта. Модель должна быть открытой: обеспечивать включение в ее состав новых подмоделей или подсистем управления (например, подсистем управления энергетическим и тепловым режимами, шихтоподготовкой и выпуском металла и т.д.).

Любую модель строят в зависимости от цели моделирования, поэтому одной из основных проблем при моделировании является проблема целевого назначения. При создании АСУ целью оптимизации может быть максимальная производительность, минимум потерь цветных металлов, снижение себестоимости продукции и некоторые другие цели.

1.4. Общие сведения о математическом моделировании

Общая классификация моделей

В основе моделирования лежит теория подобия, согласно которой абсолютное подобие может иметь место лишь при замене одного объекта другим, точно таким же. При моделировании невозможно добиться абсолютного подобия и нужно стремиться к тому, чтобы модель достаточно хорошо отражала исследуемую сторону функционирования объекта (например, процессы восстановления или осаждения металла). Поэтому в качестве одного из признаков классификации видов моделирования можно выбрать степень полноты модели и разделить модели в соответствии с этим признаком на полные, неполные и приближенные.

В основе *полного моделирования* лежит полное подобие, которое проявляется как во времени, так и в пространстве. Для *неполного моделирования* характерно неполное подобие модели изучаемому объекту («горячая» модель электролизера). *Приближенное моделирование* опирается на приближенное подобие, при котором

отдельные стороны функционирования реального объекта не моделируются совсем («холодная» модель электролизера).

В зависимости от характера изучаемых процессов в системе все виды моделирования могут быть сведены в схему, представленную на рис. 1.3.

Детерминированное моделирование отображает детерминированные процессы (т.е. процессы, в которых предполагается отсутствие всяких случайных воздействий); *стохастическое моделирование* – вероятностные процессы и события.

Статическое моделирование служит для описания поведения объекта в какой-либо момент времени (в статике). *Динамическое моделирование* отражает закон изменения состояния объекта во времени.

Дискретное моделирование служит для описания дискретных процессов. *Непрерывное моделирование* позволяет отразить непрерывные во времени процессы. Дискретно-непрерывное моделирование используют, когда необходимо выделить наличие как дискретных, так и непрерывных процессов (непрерывный процесс плавки и периодический выпуск продуктов плавки).

Линейное моделирование применяют для описания линейных процессов, т.е. процессов, связь между входными и выходными параметрами которых линейна. *Нелинейное моделирование* предназначено для описания нелинейных процессов, в которых связь между выходом и входом системы нелинейна.

В зависимости от формы представления объекта можно выделить мысленное и реальное моделирование. При *реальном моделировании* используется возможность исследования характеристик либо на реальном объекте, либо на его части. Оно подразделяется на натурное и физическое.

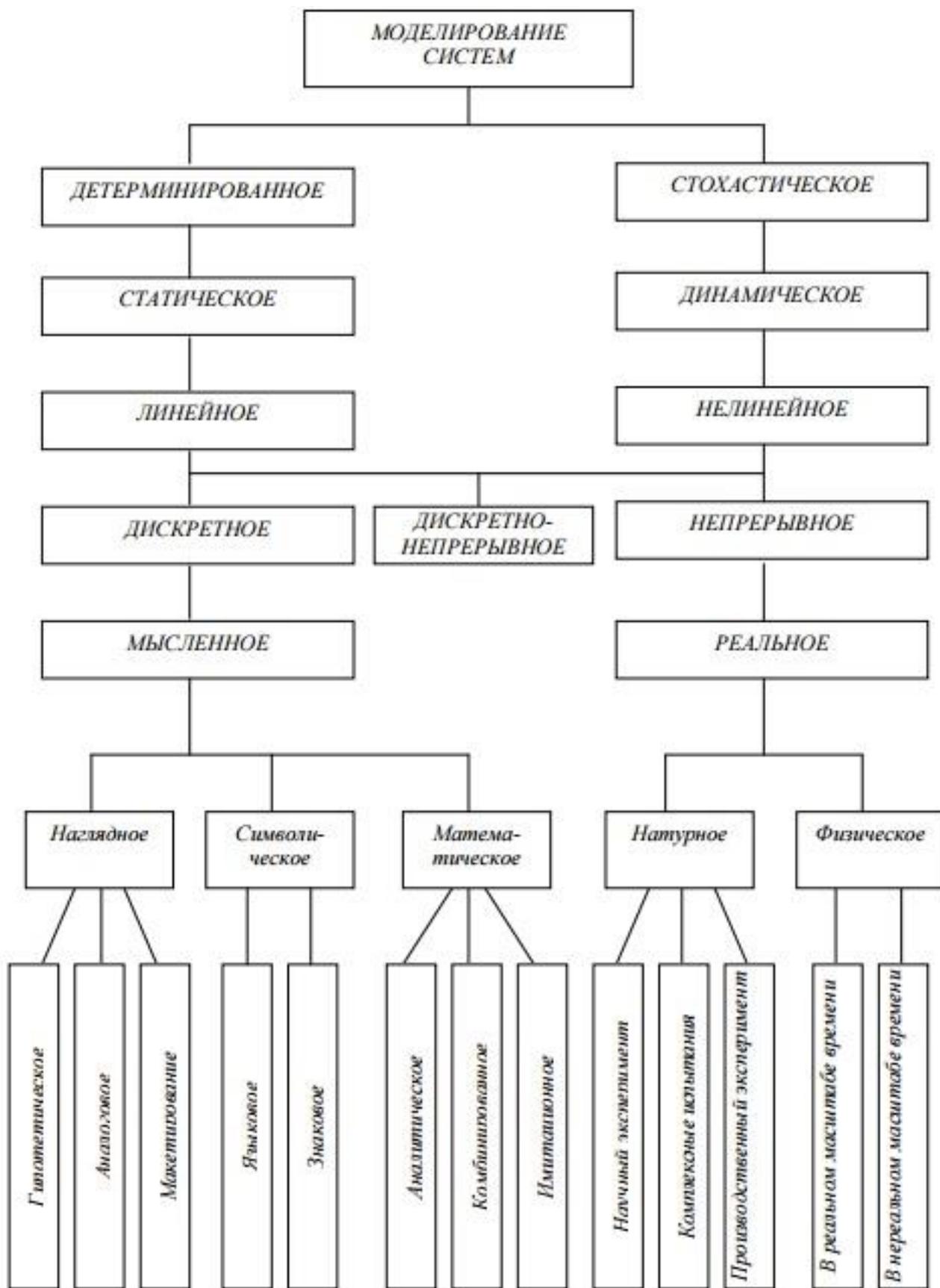


Рис. 1.3. Классификация моделей

Натурное моделирование подразделяют на научный эксперимент, комплексные испытания и производственный эксперимент, реализованные на основе теории подобия и обладающие высокой достоверностью. *Физическое моделирование* отличается от натурного тем, что исследования проводятся на установках, которые сохраняют природу явления и обладают физическим подобием в реальном и нереальном масштабе времени (моделирование процесса электролиза).

Мысленное моделирование часто является единственным способом моделирования объектов, которые либо нельзя реализовать в заданном интервале времени, либо они существуют вне условий, возможных для их физического воссоздания. Оно может быть наглядным, символическим и математическим. При *наглядном моделировании* создаются наглядные модели, которые отображают явления и процессы, протекающие в объекте. Наглядное моделирование подразделяется на гипотетическое (в основе гипотезы), аналоговое (аналогии различных уровней) и макетирование (мысленный макет). *Символическое моделирование* представляет собой искусственный процесс создания логического объекта, который замещает реальный и выражает его основные свойства (электрические схемы, чертежи, карты и т.д.). Под *математическим моделированием* понимают процесс установления соответствия данному реальному объекту некоторого математического объекта, называемого математической моделью, и исследование этой модели, позволяющее получать численные характеристики рассматриваемого реального объекта. Математическое моделирование для исследования характеристик процесса функционирования систем можно разделить на аналитическое, имитационное и комбинированное.

Для *аналитического моделирования* характерно то, что процесс функционирования элементов системы (физико-химические превращения, загрузка, выпуск, тепло- и массоперенос) описывается в виде некоторых функциональных соотношений (алгебраических, дифференциальных, конечно-разностных) или логических условий. Аналитическая модель может быть исследована различными методами:

- ◆ аналитическим, когда искомые характеристики пытаются найти в общем виде;

◆ численным, когда, не зная решения в общем виде, стремятся получить числовые результаты при конкретных начальных условиях;

◆ качественным, когда, не имея решения в явном виде, можно установить некоторые его свойства по самой модели (изменение заглубления электродов приводит к изменению токораспределения по объему ванны РТП).

Аналитический метод широко применяется для простых систем, с усложнением систем использование этого метода возможно при упрощении первоначальной модели. Численный метод по сравнению с аналитическим позволяет исследовать более широкий класс систем, но при этом полученные решения носят частный характер (процесс обжига при известном химическом составе шихты). Численный метод особенно эффективен при использовании ЭВМ.

При *имитационном моделировании* имитируются элементарные явления, составляющие процесс, с сохранением их логической структуры и последовательности протекания во времени. Основным преимуществом имитационного моделирования перед аналитическим является возможность решения более сложных задач. Имитационные модели позволяют относительно просто учитывать следующие факторы: наличие дискретных и непрерывных элементов, нелинейные характеристики элементов системы, многочисленные случайные воздействия.

Для построения имитационной модели используют машинные реализации случайных величин и функций, полученные методом *статистического моделирования* (Монте-Карло). Данный метод применяется для численного решения аналитической задачи. Особенно эффективны статистические методы при неполной информации о процессе.

Комбинированное моделирование позволяет объединить достоинства аналитического и имитационного моделирования. При этом часть подсистем описывается аналитическими моделями, а часть – имитационными.

Модели математического программирования являются классом моделей, применяемых для решения оптимизационных задач.

В самом общем случае задачи математического программирования имеют следующий вид: найти точки максимума функции

$$Z = F(\tilde{n}, \tilde{\delta})$$

при ограничениях

$$g_i(\tilde{\delta}) \leq 0, i = 1, \dots, m,$$

$$h_j(\tilde{\delta}) = 0, i = m + 1, \dots, l,$$

здесь Z – целевая функция; $\tilde{n}, \tilde{\delta}$ – векторы управляемых и неуправляемых переменных; $g_i(\tilde{\delta}), h_j(\tilde{\delta})$ – известные функции.

Для нахождения решения данной задачи в зависимости от вида структуры целевой функции и ограничений используют:

- ♦ методы линейного программирования, если $F(c, x), g_i(x), h_j(x)$ линейны относительно переменных x ;

- ♦ методы нелинейного программирования, если $F(c, x), g_i(x), h_j(x)$ нелинейны относительно переменных x ;

- ♦ методы динамического программирования, если $F(c, x)$ имеет специальную структуру;

- ♦ методы стохастического программирования, когда вектор неуправляемых переменных x случаен;

- ♦ сетевые модели (позволяют отобразить объем практических взаимообусловленных работ, последовательность их выполнения, а также логическую взаимосвязь);

- ♦ модели массового обслуживания (для решения задач обслуживания массового потребителя).

1.5. Структурно-параметрическое описание и назначение параметров объекта

Применение вычислительной техники и математических методов при проектировании сложных технических систем возможно только в том случае, если имеются их адекватные математические модели. Возможность построения иерархии моделей играет важную роль в процессе автоматизированного проектирования и создания САПР. В математической модели проектируемого объекта выделяют структурно-параметрическое описание собственно объекта и описание поведения объекта во времени и внешней среде. Таким образом, математическая модель проектируемого объекта состоит из двух частей: структурно-параметрического описания объекта с помощью набора проектных параметров и модели функционирования.

Под *структурно-параметрическим описанием ОУ* будем понимать такое его описание, которое показывает, из каких подсистем, блоков, агрегатов, деталей состоит данный объект, как эти компоненты соединены и взаимодействуют между собой, каковы их весовые, габаритные характеристики и т.д. Структурно-параметрическое описание должно давать возможность генерировать множество альтернатив ОУ, быть достаточно подробным, соответствующим этапу проектирования и доставлять информацию для моделей функционирования.

Для сложных объектов существуют различные методы структурно-параметрического описания ОУ: систематического покрытия поля, отрицания и конструирования, морфологического ящика, комбинаторного файла и т.д.

При построении модели наряду со структурным возникает потребность в параметрическом описании ОУ. Обычно такое описание дается конечным набором параметров, варьируя значения которых в определенных пределах с учетом необходимых ограничений, можно вводить в структуру ОУ различные по характеристикам подсистемы.

Параметрическое описание объекта включает в себя выделение совокупности входных переменных (внешних параметров) X_1, X_2, \dots, X_n , управляющих воздействий U_1, U_2, \dots, U_k , влияющих на процесс, выходных переменных

(зависимых параметров) Y_1, Y_2, \dots, Y_m , характеризующих протекание процесса, а также внутренних параметров модели P_1, P_2, \dots, P_l .

Управляющие воздействия U_1, U_2, \dots, U_k являются целенаправленно изменяемыми переменными и формируются на основе информации о входных переменных, которые называются управляемыми. Остальные входные переменные относятся к возмущающим воздействиям, а выходные переменные – к неуправляемым.

Внутренние параметры модели – это внутренние характеристики объекта, не зависящие от процесса моделирования, например, конструктивные параметры агрегатов, теплофизические свойства объектов и т.п.

Возмущающие воздействия и неуправляемые переменные могут быть контролируемыми (наблюдаемыми) и неконтролируемыми (ненаблюдаемыми).

В математическом моделировании используется различный математический аппарат в зависимости от характера моделируемого процесса или объекта.

Модели функционирования управляемых динамических систем описываются обыкновенными дифференциальными уравнениями. Этот класс систем характеризуется тем, что содержит модели, адекватно отражающие функционирование современных изделий машиностроения, технологических систем и т.д. Модели функционирования подразделяются:

на модели без управления

$$Y = f(t, \bar{z}, \bar{\delta}),$$

используемые для создания систем контроля, диагностики и прогнозирования технических систем, где \bar{z} – вектор фазовых состояний; t – время; $\bar{\delta}$ – набор параметров;

модели с управлением, применяемые для создания систем автоматического управления:

$$Y = f(t, \bar{z}, \bar{\delta}, \bar{u}, \mu),$$

где μ – возмущение; \bar{u} – управляющее воздействие.

Для описания процессов, протекающих в технологических агрегатах, используют математические выражения, которые составляют математическую

модель процесса или объекта управления. В зависимости от того, какие математические формулы используют для описания процессов, математические модели подразделяют на линейные и нелинейные.

Линейной называется такая математическая модель описания объекта или процесса, для построения которой используют линейные дифференциальные или другие уравнения, т.е. уравнения, в которых связь между входными и выходными параметрами является линейной.

Нелинейной называется такая математическая модель описания объекта или процесса, для построения которой используют нелинейные уравнения, т.е. уравнения, в которых связь между входными и выходными параметрами является нелинейной.

С помощью линейной модели можно описать, например, реакции, протекающие в объеме ванны металлургического агрегата, теплоперенос в стенках печного агрегата и т.д.:

$$\frac{d(V_1)}{dt} = V_1 C_A - V_2 C_A,$$

$$\frac{d(V_2)}{dt} = V_1 C_B - V_2 C_B,$$

где A, B – взаимодействующие вещества, которые имеют концентрации C_A и C_B ; V_1, V_2 – скорости изменения объема. Линейные модели могут использоваться для описания взаимосвязи входных и выходных параметров с помощью парных и множественных вероятностных (регрессионных) уравнений следующего вида:

$$T = a_0 + a_1 p;$$

$$I = b_0 + b_1 p + b_2 G + b_3 C_k + b_4 T + b_5 Y,$$

где T – температура; I – производительность; p, G, C_k, Y – технологические параметры; a, b – коэффициенты регрессионного уравнения, определенные экспериментальным путем.

При описании нелинейных моделей могут использоваться степенные зависимости x^n , показательные e^x , логарифмические, гиперболические, тригонометрические, интегральные, обратные тригонометрические и т.д.

Например, работу газа можно описать следующим выражением:

$$A = \int_{V_1}^{V_2} p dV = \int_{V_1}^{V_2} \frac{IRT}{V} dV = nRT \ln \frac{V_2}{V_1},$$

где n – число молей газа в объеме V ; R – универсальная газовая постоянная; T – абсолютная температура; V_1, V_2 – изменение объема.

С помощью нелинейных моделей можно описать сложные физико-химические процессы, протекающие в металлургических агрегатах (печь обжига, восстановительно-окислительные процессы), и процессы массопереноса и физических превращений.

Многие взаимосвязи между входными и выходными параметрами могут быть описаны с помощью вероятностных нелинейных уравнений, например, вида

$$\eta_0 = a_0 + a_1 T + a_2 T^2 + a_3 T^3,$$

где η_T – выход по току; a_i – коэффициенты регрессии, значения которых определяются экспериментально; T – температура электролита.

1.6. Дискретные и непрерывные модели

При построении моделей сложных систем большое значение имеет идея прерывности, на основе которой сложные явления поддаются описанию как закономерно составленные из простых частей. Подчеркивая преобладающую роль принципа прерывности, нельзя игнорировать непрерывность модели, где находит отражение одно из важнейших свойств системности, заключающееся в том, что система не есть простая сумма составляющих ее элементов.

Идея прерывности применяется при описании процессов с сосредоточенными и распределенными параметрами.

Используя принципы соответствия, конкретности и прерывности для систем управления отдельными агрегатами, цехами и заводами, можно построить расчленённые модели, обозримые как в целом, так и для каждой из структурных единиц. Чем сложнее система, тем более крупную структурную единицу нужно рассматривать на каждой ступени иерархии, а модель всей системы можно представить как систему подмоделей. Например, децентрализованная систем управления процессом электролиза включает управление электролизерами на базе

микропроцессоров и управление корпусом путем изменения установок электролизеров.

При этом между структурными единицами нужно оставить лишь минимум необходимых связей, учитывая, что внутри каждой единицы существует саморегулирование. Например, при децентрализованной системе управления электролизерами ЭВМ верхнего уровня освобождается от управления межполюсным расстоянием (МПР) и питанием электролизеров глиноземом, что возлагается на микропроцессорные системы управления. При этом в функции ЭВМ верхнего уровня остается координация работы микропроцессоров.

Контрольные вопросы и задания

1. Что понимается под объектом моделирования?
2. Что такое гипотеза в моделировании?
3. Дайте определение модели.
4. Что такое математическая модель?
5. Приведите пример аналогии в физических процессах.
6. Дайте классификацию процессов как объектов моделирования.
7. Чем отличаются стохастические процессы от детерминированных?
8. Опишите постановку задачи моделирования в общем виде.
9. Дайте общую классификацию математических моделей.
10. Какова структура модели математического программирования?
11. Что понимают под структурно-параметрическим описанием объекта моделирования?
12. В чем состоит различие между линейными и нелинейными моделями?
13. В каких случаях используется корреляционный коэффициент, а в каких - корреляционное отношение как критерий адекватности модели?
14. Дайте классификацию моделируемых процессов по характеру их протекания.

2. МАТЕМАТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ И ОПТИМИЗАЦИИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ В МАШИНОСТРОЕНИИ

2.1. Основные понятия и определения

Из множества задач, решаемых технологами при исследованиях существующих процессов и создании новых, можно выделить три весьма распространенных вида [1, 3, 5, 11]:

- выявление количественных зависимостей между параметрами процесса;
- отыскание оптимальных условий протекания процесса;
- выбор оптимального состава многокомпонентных смесей.

Так как в большинстве случаев информация о закономерностях взаимосвязи технологических параметров чрезвычайно мала, для решения поставленных задач целесообразно использование кибернетического подхода, в основе которого лежит идея «черного ящика». Он представляет собой систему связей, недоступную для наблюдения, так как о механизме процесса нам ничего неизвестно или известно лишь частично [1, 2, 5, 8, 13] (рис. 2.1).

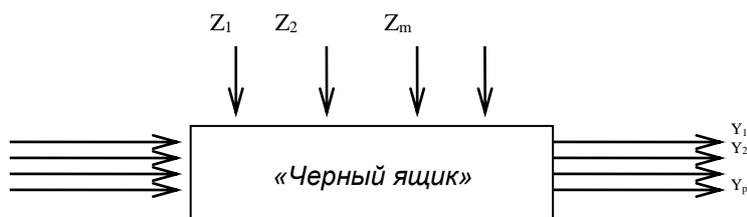


Рис. 2.1. Схема «черного» ящика

Входные параметры называются факторами процесса и могут быть управляемыми x и неуправляемыми Z . Выходные параметры Y называются откликами и являются результатом экспериментов. Зависимость между выходным параметром-откликом и входными параметрами-факторами называется функцией отклика и имеет общий вид

$$Y = f(x_1, x_2, x_3, \dots, x_k). \quad (2.1)$$

Уравнению (2.1) соответствует некоторая гиперповерхность в многомерном пространстве, называемая поверхностью отклика, а пространство, в котором существует указанная поверхность, называется факторным пространством. Если фактор один, то поверхность вырождается в кривую, если факторов два, это будет «обычная» поверхность, наконец, если факторов три и больше, мы имеем дело с гиперповерхностью, которую нельзя изобразить, но с которой, тем не менее, можно работать, как с «обычной» поверхностью.

На рис. 2.2 пунктирными линиями показана некоторая поверхность отклика в двухфакторном пространстве кодированных факторов. Пунктирной линией подчеркивается то, что эта поверхность существует, но неизвестна нам. Информацию об этой поверхности можно получить в виде любого числа принадлежащих ей точек путем проведения эксперимента при различных сочетаниях факторов x_1 и x_2 . Полученная информация используется для построения аппроксимирующей поверхности (сплошная линия на рис. 2.2), которая должна проходить как можно ближе к поверхности отклика. При обработке эмпирической информации, в которой содержатся случайные ошибки, с целью получения аппроксимирующей зависимости, описывающей функцию отклика, широко применяется метод наименьших квадратов. В методе наименьших квадратов в качестве критерия точности аппроксимации используется следующий показатель: сумма квадратов отклонений между известными значениями исследуемой функции отклика и соответствующими значениями аппроксимирующей функции должна быть минимальной:

$$\sum_{u=2}^N (Y_u - \hat{Y})^2 = \min , \quad (2.2)$$

где N – число экспериментальных точек, по которым проводится аппроксимация,

u – номер точки;

знак $\hat{}$ – обозначения, предсказанные аппроксимирующим выражением.

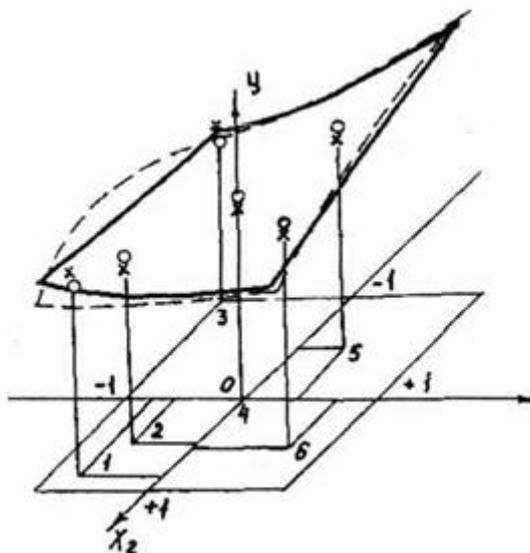


Рис. 2.2. Поверхность отклика (пунктир)
и аппроксимирующая поверхность (сплошные линии)

В математическом планировании эксперимента наиболее широкое распространение в качестве аппроксимирующей функции получили полиномы некоторой степени:

$$Y = b_0 + \sum_{i=1}^k b_i x_i + \sum_{\substack{i=1 \\ j=1}}^k b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2 + \dots, \quad (2.3)$$

где b_0 , b_i , b_{ij} , b_{ii} , ... – выборочные коэффициенты регрессии, которые можно получить, пользуясь результатами эксперимента.

Полученное эмпирическим путем уравнение регрессии часто называют математической моделью процесса.

Так как степень полинома заранее предсказать нельзя, то пользуются идеей шагового поиска, т.е. сначала процесс описывается линейной моделью, и если качество модели неудовлетворительное, то увеличивают число членов полинома, повышая его степень.

Для получения выборочных оценок, коэффициентов уравнения регрессии можно организовать проведение эксперимента двумя принципиально различными путями.

Информация, собранная при так называемом «активном» эксперименте по математически обоснованному плану, учитывающему цели эксперимента и методы обработки его результатов, имеет много большую ценность, чем информация от

эксперимента, поставленного по традиционной методике, когда изменяется каждый фактор в отдельности (так называемый «пассивный» эксперимент). При этом практически всегда уменьшаются затраты ресурсов (материальных и временных) на активный эксперимент и существенно облегчается интерпретация моделей. В работах В.В. Налимова [11, 19, 22] на основе большого опыта сформулированы преимущества активных экспериментов:

- минимизируется число опытов;
- оптимально используется факторное пространство;
- вводится четкая логика для всех процедур, последовательно совершаемых экспериментатором;
- рандомизируются условия опыта;
- благодаря направленной организации эксперимента выполняются исходные предпосылки статистического анализа;
- оценивается элемент неопределенности, дающий возможность сопоставлять результаты, полученные разными исследованиями;
- повышается контроль за точностью эксперимента.

Спланировать эксперимент – это, в частности, значит:

- а) выбрать из N_x действующих в технологическом процессе факторов те наиболее существенные факторы $x_1, x_2, x_3, \dots, x_i, \dots, x_k$, уровни которых технолог собираются изменить;
- б) наметить пределы их варьирования $(x_i)_{\min}$ и $(x_i)_{\max}$;
- в) наметить комбинацию уровней k -факторов x_i , при которых будет исследоваться процесс;
- г) определить число повторений опытов и измерений в каждой из выбранных комбинаций k -факторов.

Началом экспериментального исследования является изучение и анализ всех имеющихся данных о технологическом процессе. В соответствии с идеей шагового поиска эксперимент проводится в несколько этапов, число которых и действия на каждом из них зависят от конечной цели исследования. Из всего многообразия конечных целей исследования можно выделить две:

1) найти адекватное описание функции отклика в заданной части факторного пространства;

2) найти оптимальные условия протекания процесса. В зависимости от сочетания результата предыдущего этапа и вида конечной цели принимается решение о действиях на следующих этапах исследования.

При отыскании оптимальных режимов технологических процессов, в случае отсутствия данных об области этих режимов, целесообразно применять шаговые или последовательные методы планирования, такие как методы крутого восхождения или последовательного симплекса планирования. В этих методах поверхность отклика аппроксимируется в локальной области, достаточно далекой от экстремума, линейным приближением с помощью минимально необходимого числа экспериментальных точек. Затем продвижение в область экстремума идет в направлении градиента линейного приближения или путем отбрасывания вершин симплекса с минимальным значением отклика. Эти методы факторного планирования эксперимента позволяют достигнуть «почти стационарной области» – области оптимальных режимов технологического процесса. Если эта область достигнута вышеуказанными методами или получена на основании анализа опытных данных, необходимо для выяснения взаимосвязи отклика и варьируемых параметров построить математическую модель исследуемого процесса, уравнение связи. Обычно в области экстремума поверхность отклика аппроксимируется полиномами второго и более высших порядков.

2.2. Определение и выбор параметров оптимизации и факторов технологических процессов

При проведении «активного» эксперимента, т.е. по методике математического планирования эксперимента, к отклику и факторам исследуемого технологического процесса предъявляется ряд принципиальных требований [3, 5, 9, 12].

Отклик – это результат опыта в соответствующих условиях. Его часто называют также функцией цели, критерием эффективности, критерием

оптимальности, параметром оптимизации и др. Все приведенные термины можно считать синонимами.

Отклики должны отвечать следующим требованиям:

- а) характеризовать наиболее общие свойства технологического процесса;
- б) оцениваться количественно и быть однозначными, причем важно, чтобы они имели физический смысл и легко вычислялись;
- в) обладать статистической эффективностью, т.е. быть нечувствительными к малым случайным воздействиям, и иметь минимальную ошибку воспроизводимости для параллельных опытов одной серии.

В случае, если интересующий нас признак качественный, а критерий, по которому мы оцениваем качество – сложный, необходимо использовать ранговый подход. Ранг – это количественная оценка параметра оптимизации, но она носит условный (субъективный) характер. Мы ставим в соответствие качественному признаку некоторое число – ранг.

Желательно, чтобы параметров оптимизации было как можно меньше. Однако не следует добиваться уменьшения числа параметров оптимизации за счет полноты характеристики процесса. При планировании эксперимента целесообразно измерять все параметры, затем оценивать корреляцию между ними и строить модели для их минимально возможного числа или же воспользоваться обобщенным параметром. Построение обобщенного параметра оптимизации связано с созданием единого признака, количественно определяющего закономерности технологического процесса со многими входными параметрами. В работе рассмотрено несколько различных способов построения обобщенного показателя.

Важным вопросом при планировании эксперимента является определение и выбор входящих параметров – факторов технологического процесса.

Факторы могут быть качественными и количественными. В инженерной практике, как правило, встречаются количественные факторы: температура, давление, концентрация и т.п. К качественным факторам относятся: вид катализатора, тип аппарата, исполнители и др. Хотя качественным факторам не соответствует числовая шкала в том смысле, как это понимается для

количественных факторов, можно построить условную порядковую шкалу, в которой уровням качественного фактора соответствуют числа натурального ряда, т.е. производится кодирование.

Каждый фактор, участвующий в процессе, имеет определенный предел изменения своей величины. Совокупность всех значений, которые принимает фактор, называется областью определения фактора. Но в области определения необходимо найти локальную подобласть для планирования эксперимента, т.е. для каждого фактора необходимо указать интервал варьирования параметров, в пределах которого проводятся исследования. Конкретные значения, которые имеет фактор во время эксперимента, называются уровнями.

Факторы в области определения исследуемых процессов должны отвечать следующим принципиальным требованиям. Они должны быть:

- а) управляемыми, т.е. в течение всего опыта экспериментатор может управлять каждым фактором в отдельности;
- б) совместными, т.е. все комбинации уровней осуществимы и безопасны;
- в) независимыми, т.е. необходимо обеспечить возможность их поддержания на любом уровне в течение всего эксперимента вне зависимости от уровней других факторов.

В практических задачах области определения факторов, как правило, ограничены. Ограничения могут носить принципиальный либо технический характер.

В результате сбора предварительной априорной информации об изучаемом процессе исследователь должен составить полный список факторов, исходя из того, что лучше назвать несколько малозначимых факторов, чем пропустить один существенный, а также задать ориентировочные пределы изменения факторов с учетом требований к ним. Если количество факторов окажется большим (семь и более), необходимо обратиться к методам отсеивания незначимых факторов.

В качестве примера рассмотрим порядок выбора основных факторов и области их определения при изучении процесса диффузионной сварки в вакууме (ДСВ) жаропрочного никелевого сплава ВЖЛ12У с высокопрочной сталью ЭИ961.

Технологический процесс диффузионной сварки изучался применительно к получению сварного соединения рабочего колеса с валом роторов турбин малоразмерных ГТД.

Обычно для количественной оценки качества сварных изделий используют показатели механических свойств полученных соединений. Испытание на растяжение является основным и наиболее распространенным методом исследования механических свойств материалов. Использование этого метода для контрольных испытаний регламентируется Государственным стандартом. Важнейшей характеристикой свойств материалов, а также соединений, полученных сваркой, является предел прочности σ_b (временное сопротивление), определяемый при испытаниях на растяжение.

Следует отметить, что этот вид испытаний не требует больших затрат времени и средств, чем, например, усталостные испытания. Учитывая, что технология ДСВ стали ЭИ961 и сплава ВЖЛ12У предусматривает внесение в стык мягкой прослойки из чистого никеля, то для ряда испытаний, например, определения ударной вязкости этого соединения, требуется уточнение и доработка стандартных методик. Предел прочности при растяжении отвечает всем требованиям, предъявляемым к критериям оптимизации. В дальнейшем для уточнения механических свойств сварных узлов была проведена серия испытаний – высокотемпературных, усталостных, комплексных на специальных стендах и натуральных изделиях.

При диффузионной сварке соединяемые детали пластически деформируются под воздействием давления сжатия и температуры сварки. Величина остаточной деформации в значительной мере определяет доработку узла после сварки, а также необходимые припуски на детали до сварки.

Выбор величины остаточной деформации в качестве критерия оптимизации позволил выбрать режим ДСВ роторов, при котором пластическая деформация ограничивается минимально необходимой.

Следовательно, в качестве откликов или параметров оптимизации целесообразно выбрать прочность на растяжение Y и остаточную макропластическую деформацию сварного соединения Z . Причем целью

исследований является определение оптимальных значений технологических факторов, при которых сварное соединение будет иметь максимальную прочность Y_{\max} при минимально возможной остаточной пластической деформации Z_{\min} .

Процесс диффузионной сварки в вакууме характеризуется целым рядом факторов. К числу наиболее ответственных технологических параметров следует отнести: температуру сварки; удельное давление сжатия; время сварки; глубину вакуума; класс шероховатости свариваемых поверхностей; материал и геометрию прослой; род растворителей для удаления загрязнений с поверхности перед сваркой; характер термической и механической обработки сплавов перед сваркой.

Анализ технологических параметров показывает, что их количество достигает десяти и среди них имеются такие, которые поддаются количественной оценке или отражают качественные характеристики технологии ДСВ.

Исходя из требований, предъявляемых к факторам, и учитывая, что с увеличением числа факторов при оптимальном планировании экспериментов значительно увеличивается объем математической обработки результатов, целесообразно при исследовании процесса ДСВ жаропрочных сплавов на основании опыта проведенных работ выбрать минимально возможное число варьируемых переменных.

Для отсеивания факторов и выбора наиболее ответственных из них в данной работе были проанализированы результаты исследований по ДСВ различных сочетаний материалов, а также проведены предварительные эксперименты.

Анализ литературных данных показал, что такие качественные факторы, как род растворителей для удаления загрязнений и характер термической обработки сплавов перед сваркой, незначительно зависят от сочетания металлических сплавов и могут быть выбраны предварительно. Так, в качестве растворителей для удаления загрязнений можно выбрать бензин и ацетон. Исследуемые сплавы должны проходить термическую обработку согласно техническим условиям на поставку.

Понижение класса поверхности, т.е. увеличение высоты микронеровностей, как правило, приводит к уменьшению площади фактического контакта, на которой развиваются диффузионные процессы, а следовательно, и снижению

конструкционной прочности соединения. Особенно это свойственно материалам, поверхностный слой которых характеризуется высокой твердостью (ВЖЛ12У – $H_{\mu} = 5500$ МПа, ЭИ961 – $H_{\mu} = 3200$ МПа), а также материалам, которые подвергаются ДСВ на режимах с пониженными значениями температуры сварки. Применяемый в данной технологии сварки прослой из чистого никеля значительно снижает влияние микрогеометрии поверхности на величину площади фактического контакта, так как при температурах сварки он достаточно пластичен ($\sigma_T \leq 30$ МПа).

Следует отметить, что с повышением класса чистоты обработки поверхности условия формирования площади фактического контакта улучшаются. В связи с тем, что получение класса шероховатости поверхности выше восьмого связано с усложнением технологии, а высота микронеровностей при 4-5-м классах сравнима с толщиной прослоя (50 мкм), целесообразно использовать детали роторов со шлифованными до седьмого класса поверхностями под сварку.

На основании опыта работ и анализа свариваемых материалов (сплав ВЖЛ12У и сталь ЭИ961), а также предварительных экспериментов были выбраны материал и толщина прослоя. Как показала расчетная оценка, мягкая прослойка из чистого никеля толщиной 0,05 мм на прочность натуральных узлов не оказывает существенного влияния.

Одним из параметров технологического процесса ДСВ жаропрочных сплавов является глубина вакуума. В настоящей технологии для разрушения слоя окислов на свариваемых поверхностях, а также для сохранения чистых поверхностей в течение всего времени сварки нагрев деталей осуществляется в вакууме 10^{-4} – 10^{-5} мм рт. ст., обычно применяемом для ДСВ на серийных сварных установках. Как показывают эксперименты, при нагреве выше 873 К в вакууме 10^{-4} – 10^{-5} мм рт. ст. происходит разрушение окисных пленок на свариваемых материалах, причем более интенсивно на стали ЭИ961.

Таким образом, шероховатость поверхности, материал, геометрия прослоя и глубина вакуума могут быть приняты за постоянные параметры, являющиеся оптимальными. Такие факторы, как температура сварки, удельное давление сжатия и время сварки, являются основными параметрами процесса ДСВ, определяющими

процессы формирования физического контакта, диффузионного массообмена в зоне сварки, структурные превращения основных материалов, а следовательно, и качество соединения в целом. Эти параметры полностью отвечают требованиям, предъявляемым к факторам процессов, исследуемых с помощью методики оптимального планирования эксперимента.

Температурный режим диффузионной сварки в вакууме влияет на все стадии процесса, и при выборе интервала варьирования температуры сварки $T_{\text{св}}$ необходимо это учитывать.

Удаление окисных пленок с поверхности высоколегированных сталей и сплавов, как правило, наиболее интенсивно происходит при температурах, превышающих 1223–1273 К в вакууме 10^{-4} – 10^{-5} мм рт. ст., что и подтверждают эксперименты на изучаемых материалах. Следовательно, в целях сокращения времени, необходимого для разрушения окисных пленок в условиях сварки, целесообразно, чтобы нижняя граница интервала варьирования $T_{\text{св}}$ была не ниже 1273 К.

Развитие физического контакта в значительной степени определяется механическими свойствами поверхностных слоев соединяемых деталей. При температурах выше 1273 К пластичность изучаемых сплавов резко увеличивается. Это позволяет при сравнительно небольших усилиях сжатия обеспечить физический контакт через прослой чистого никеля по всей поверхности соединения.

Температурный режим ДСВ стимулирует как диффузионные процессы в самих швах, так и взаимную диффузию в переходной зоне между ними. Исследуемые материалы являются жаропрочными высоколегированными сплавами. Наличие в структуре упрочняющих фаз, например, в сплаве ВЖЛ12У γ -фазы, ограничивает температуру нагрева ввиду возможных необратимых структурных изменений.

Более чувствительным к перегреву является сплав ВЖЛ12У. Так, нагрев выше температуры закалки 1473 К приводит к необратимым структурным изменениям, вследствие чего прочность и жаростойкость этого материала значительно снижаются.

Наличие в сплавах большого количества легирующих элементов, особенно в ВЖЛ 12У, затрудняет протекание диффузионных процессов, а также процессов релаксации внутренних напряжений. Так, для снятия внутренних напряжений в сплаве ВЖЛ12У проводится отжиг при 1223 К в течение двух часов. Следовательно, для стимулирования взаимной диффузии свариваемых сплавов целесообразен нагрев до температур не ниже 1273 К. Следует отметить, что уменьшение температуры сварки ниже 1273 К, как показал опыт, приводит к резкому падению прочности соединения.

Таким образом, диапазон варьирования сварочных температур целесообразно принять равным от 1273 до 1473 К, чтобы полностью отключить возможность перегрева сплава ВЖЛ12У.

Удельное давление сжатия для жаропрочных сплавов обычно выбирается экспериментально. Величина удельного давления влияет, в первую очередь, на остаточную деформацию сварного соединения, которую целесообразно ограничить. Минимальные макропластические деформации сварного соединения могут быть получены лишь при сварке с применением относительно низких удельных давлений. В то же время давление сжатия должно быть достаточным для образования физического контакта по всей поверхности соединяемых деталей, что определяется пластическими свойствами используемых материалов. В связи с тем, что для свариваемых сплавов точные характеристики пластичности при высоких температурах отсутствуют, интервал варьирования удельного давления сжатия выбран достаточно широким: от 5 до 20 МПа.

В настоящее время длительность изотермической выдержки при сварке, особенно для разнородных сложных сочетаний, выбирается экспериментально. Длительные выдержки при сварке значительно снижают производительность процесса ДСВ, а также приводят к увеличению остаточных деформаций сварного соединения. Ограничение времени может привести к незавершенности формирования физического контакта и снижению величины зоны объемного взаимодействия между свариваемыми материалами. Предварительные эксперименты показали, что выдержка в течение 15–20 мин. в диапазоне вышеуказанных интервалов температур и удельных давлений обеспечивает удовлетворительное качество соединения; дальнейшее повышение длительности

сварки до 1 ч не приводит к заметному росту механических свойств сварных образцов.

Область факторного пространства, в которой целесообразно исследовать процесс ДСВ сплава ВЖЛ12У со сталью ЭИ961, следующая:

$$\begin{aligned} 1273 \text{ К} &\leq T_{\text{св}} \leq 1473 \text{ К}; \\ 5 \text{ МПа} &\leq P_{\text{св}} \leq 20 \text{ МПа}; \\ 5 \text{ мин} &\leq \tau_{\text{св}} \leq 20 \text{ мин}. \end{aligned} \quad (2.4)$$

Таким образом, анализ информации о процессе диффузионной сварки и свойствах исследуемых сплавов, а также небольшое количество предварительных экспериментов позволили существенно уменьшить число варьируемых факторов и достаточно объективно определить границы изучаемого факторного пространства.

В табл. 2.1 представлены наиболее важные факторы и параметры оптимизации для ряда новых технологических процессов изготовления деталей летательных аппаратов и их сборки. При определении факторов и параметров оптимизации пользовались описанной выше схемой их выбора на основании теоретического и экспериментального опыта по исследованию рассматриваемых процессов.

Контрольные вопросы

1. В чем заключается сущность принятого в планировании эксперимента подхода, основанного на идее «черного ящика»?
2. Понятия факторов и откликов исследуемых технологических процессов.
3. Какой метод лежит в основе получения аппроксимирующей зависимости, описывающей функцию отклика?
4. Каковы преимущества методики математического планирования эксперимента?
5. Что означает выражение «планировать эксперимент»?
6. Какие требования предъявляются к откликам исследуемого технологического процесса?
7. Какие требования предъявляются к факторам исследуемого технологического процесса?
8. В чем заключается порядок выбора основных факторов и области их варьирования при изучении технологического процесса?

**Основные факторы и критерии оптимизации (отклики) для некоторых технологических процессов
в производстве летательных аппаратов**

Технол. процесс	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Y_1	Y_2	Y_3
Диффузионная сварка и пайка	Температура	Давление сжатия	Время	Глубина вакуума	Химич. состав прослой (припоя)	Толщина прослой (припоя)	Шероховатость поверхности	Прочность	Пластичность	Деформация
Магнитоимпульсная обработка материалов	Внешний радиус индуктора	Внутренний радиус индуктора	Длина индуктора	Число витков	Радиус детали	Толщина детали	Напряжение батареи	Перемещение заготовки	Скорость перемещения заготовки	—
Ионно-плазменное напыление	Ток разряда	Индукция магнитного поля	Потенциал подложки	Давление реакционно-го газа	Энергия ионов	Плотность потока энергии	Температура	Адгезия	Износостойкость	Остаточные напряжения
Плазменное (газотермическое) напыление	Сила тока дуги	Расход плазмособраз. газа	Расход водорода	Дистанция напыления	Скорость плазматрона	Шаг попереч. перемещ.	Расход порошка	Адгезия	Износостойкость	Пористость
Штамповка эластичной средой	Высота пуансона матрицы	Величина припуска	Толщина заготовки	Твердость эластичной среды	Механич. свойства материалов	Величина противодействия	Форма инструмента	Угол пружинения	Отклонение геометр. детали	Угол скола
Дорнование и запрессовка болтов	Диаметр отверстия	Угол заборного конуса	Натяг	Скорость	Пластичность материала	Чувств. к динам. нагрузкам	Вид смазки	Усилие дорнования	Точность отверстий	Предел выносливости
Клепка	Диаметр отверстия	Толщина пакета	Схема клепки	Тип заклепки	Усиление расклепывания	Скорость деформирования	—	Натяг	Герметичность	Ресурс

3. РАЦИОНАЛЬНОЕ ПЛАНИРОВАНИЕ

Рациональное планирование экспериментов позволяет при минимальном числе опытов наиболее равномерно охватить всю площадь таблицы возможных сочетаний влияющих факторов. В этом случае эксперимент планируется так, чтобы ни в одной строке и ни в одном столбце не было повторных сочетаний. На рис. 3.1 показан один из возможных планов такого сочетания четырех факторов, каждый из которых может принимать пять значений.

Номер столбца средних (по значению) квадратов соответствует номеру уровня фактора x_1 , а номер строки средних квадратов – номеру уровня фактора x_3 . Из 25 возможных сочетаний факторов x_2 и x_4 в каждом из средних квадратов мы выбираем только одно, обозначенное зачерненной клеткой, причем в каждой строке и в каждом столбце мелких квадратов должна быть только одна такая клетка. Нетрудно убедиться, что для каждого уровня одного из факторов, например для $x_1 = 1$, все уровни прочих факторов встречаются одинаково часто. Так, в этом случае: $x_2 = 3, 4, 5, 2, 1$; $x_3 = 1, 2, 3, 4, 5$ и $x_4 = 1, 2, 3, 5, 4$. Поэтому при определении результатов для $x_1 = 1$ влияние трех других факторов усреднится и результат будет соответствовать $x_2 = 3$; $x_3 = 3$; $x_4 = 3$.

Производя такое усреднение для каждого уровня фактора x_1 , можно найти зависимость результата только от этого фактора при нейтрализации влияния остальных трех факторов. Аналогично можно выявить влияние только фактора x_2 при нейтрализации x_1 , x_3 и x_4 . Меняя порядок усреднения, можно из одних и тех же данных 25 опытов найти влияние всех четырех первичных факторов. Таким образом, данная методика позволяет заменить полное число сочетаний влияющих факторов, равное 625, всего лишь 25 специально подобранными сочетаниями факторов, т.е. сократить объем экспериментов в 25 раз.

		1					2					3					4					5				
		x1																								
x3	x2	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	x4																									
1	1			■																						
	2																									
	3																									
	4																									
	5																									
2	1																									
	2																									
	3																									
	4																									
	5																									
3	1																									
	2																									
	3																									
	4																									
	5																									
4	1																									
	2																									
	3																									
	4																									
	5																									
5	1																									
	2																									
	3																									
	4																									
	5																									

Рис. 3.1. Рациональный план эксперимента для четырех факторов и пяти уровней

Методика построения комбинационных квадратов. Весь последующий анализ проводится для четырех первичных независимых друг от друга факторов. При этом решение более простых случаев зависимости результатов от трех или двух факторов может быть получено из основного случая при условии, что один или два фактора будут постоянными.

Строится большой комбинационный квадрат (рис. 3.2) и рядом помещается средний квадрат в окружении четырех таких же средних квадратов, примыкающих к нему крест-накрест (см. верхнюю часть рис. 3.2). Пронумеруем в среднем квадрате все клетки от 1 до 25. Центральную клетку в большом квадрате обозначим цифрой 13, т.е. цифрой, располагающейся в центральной клетке среднего квадрата.

Если использовать этот же прием, но отсчет вести не от центральной клетки 13, а от какой-либо другой, например 19, то придется продолжить диагональ в квадраты, примыкающие к отдельному среднему квадрату, т.е. взять клетки 2, 23, 19, 15, 6. При переносе этих клеток в большой квадрат они расположатся вдоль ломаной линии в четвертой строке большого квадрата. Продолжая это построение, получаем расположение всех 25 клеток в большом комбинационном квадрате, причем все клетки будут иметь различные номера, т.е. соответствовать различным сочетаниям первичных факторов.

Аналогичными приемами могут быть построены комбинационные квадраты для 7-го, 8-го и т.д. уровней каждого из четырех факторов.

Методика обработки данных. Сущность методики рассмотрим на примере для четырех факторов x_1, x_2, x_3, x_4 , каждый из которых может принимать одно из пяти следующих значений: 1, 2, 3, 4, 5. Опыты выполнены по плану, изображенному на рис. 3.3. В клетках, обозначающих комбинации уровней факторов в опытах, записаны значения переменной величины y , полученные в результате проведения эксперимента. Данные опытов сгруппируем по значениям факторов (табл. 3.1, 3.2).

Таблица 3.1

Результаты эксперимента

x_2	x_1					Среднее
	1	2	3	4	5	
1	29	15	33	24	49	30
2	23	34	42	28	33	32
3	22	43	36	32	37	34
4	26	37	20	56	41	36
5	30	21	39	50	50	38
Среднее	26	30	34	38	42	—

Результаты эксперимента

x_4	x_3					Среднее
	1	2	3	4	5	
1	22	39	34	49	56	40
2	33	26	33	50	43	37
3	24	37	30	37	42	34
4	21	28	41	35	29	31
5	20	15	32	23	50	28
Среднее	24	29	34	39	44	—

Для каждого уровня первого, второго, третьего и четвертого факторов находим средние значения величины y . Нанося в системе координат средние значения величины y , соответствующие уровням фактора x_1 , получаем график зависимости y от x_1 . Аналогично строятся графики зависимостей y от остальных факторов.

Согласно полученным данным функция y линейно зависит от каждого из факторов. Параметры этих зависимостей найдем с использованием метода наименьших квадратов. Так как все графики частных зависимостей величины y от каждого из факторов аппроксимируются с достаточной точностью прямыми, то зависимость y от всех факторов может быть представлена суммой частных зависимостей.

Свободный член этой зависимости b_0 определяется следующим образом. Подставляя в найденное уравнение значения факторов первого опыта и величину y_1 , полученную в этом опыте, находим значение b_{01} . Аналогично находим для последующих опытов величины $b_{01}, b_{02}, \dots, b_{0k}$. Искомый параметр b_0 определяем как среднеарифметическое величин $b_{01}, b_{02}, \dots, b_{0k}$, где k – число проведенных опытов.

Таким образом, методика рационального планирования эксперимента, существенно уменьшая число необходимых опытов, позволяет исследовать некоторые многофакторные системы, однако она имеет и ряд недостатков. К числу основных недостатков относятся: отсутствие статистического

обоснования результатов, которое может привести к ошибочным выводам; ограниченная область применения (описанная методика построения плана эксперимента пригодна не для всякого числа уровней факторов); несовершенство и недостаточная обоснованность способа получения эмпирических зависимостей.

Эти недостатки устраняются в методах статистического планирования эксперимента, рассмотренных в следующих параграфах.

Контрольные вопросы

1. Что такое рациональное планирование?
2. Опишите методику построения комбинационных квадратов.
3. Как проводится обработка экспериментальных данных при использовании рационального планирования?

4. КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ ОДНОФАКТОРНЫЙ АНАЛИЗ

При исследовании технологических процессов часто приходится иметь дело с *взаимосвязанными* показателями. При этом часто связь, существующая между двумя или несколькими показателями, затушевывается, усложняется наложением действия других причин (факторов). Изучить, насколько изменение одного показателя зависит от изменения другого (или нескольких) – одна из важнейших задач статистики.

Следует различать функциональные и корреляционные связи. При **функциональной** зависимости каждому значению одной переменной строго соответствует определенное значение другой переменной. Зависимость, при которой одному значению переменной (x) может соответствовать (в силу наложения действия других причин) множество значений другой переменной (y), называют **корреляционной**. Корреляционная зависимость проявляется лишь на основе массового наблюдения.

Примером корреляционной зависимости может служить зависимость производительности труда от стажа работников, зависимость урожайности от срока сева, зависимость годового удоя коров от количества отелов и т.п. Наиболее простым случаем корреляционной зависимости является *парная* корреляция, т.е. зависимость между двумя признаками (результативным и одним из факторных).

Основными задачами при изучении корреляционной зависимости являются:

- 1) отыскание формы связи в виде математической формулы, выражающей эту зависимость, y от x ;
- 2) измерение тесноты такой зависимости.

Решение первой задачи, т.е. определение формы связи с последующим отысканием параметров уравнения, называется нахождением уравнения связи

(уравнения регрессии). Результативный показатель, рассматриваемый как функция x , обозначается y_x (читается: «игрек, выровненный по икс»).

Возможны различные формы связи:

1) прямолинейная: $\bar{y}_x = a_0 + a_1 x$;

2) криволинейная в виде:

а) параболы второго порядка: $\bar{y}_x = a_0 + a_1 x + a_2 x^2$ (или высших порядков);

б) гиперболы: $\bar{y}_x = a_0 + \frac{a_1}{x} + a_0 x^x$;

в) показательной функции: $\bar{y}_x = a_0 a_1^x$.

Параметры для всех уравнений связи чаще всего определяют из так называемой системы нормальных уравнений, отвечающих требованию метода наименьших квадратов (МНК). Это требование можно записать как $\sum (y - \bar{y}_x)^2 \rightarrow \min$ или при линейной зависимости ($\bar{y}_x = a_0 + a_1 x$) как $\sum (y - a_0 - a_1 x)^2 \rightarrow \min$, т.е. требуется определить, при каких значениях параметров a_0 и a_1 сумма квадратов отклонений y от \bar{y}_x будет минимальной. Найдя частные производные указанной суммы по a_0 и a_1 и приравняв их нулю, легко записать систему уравнений, решение которой и дает параметры искомой функции, т.е. уравнения регрессии.

Так, система нормальных уравнений при линейной зависимости имеет вид:

$$\begin{cases} na_0 + a_1 \sum x = \sum y; \\ a_0 \sum x + a_1 \sum x^2 = \sum xy. \end{cases} \quad (4.1)$$

Необходимые для решения данной системы показатели $n, \sum x, \sum y, \sum x^2, \sum xy$ определяются по наблюдаемым эмпирическим данным (см. задачу А). Решить такую систему нормальных уравнений можно разными способами.

Для нахождения a_0 и a_1 и при линейной зависимости можно воспользоваться формулами

$$a_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \text{ или } a_1 = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}. \quad (4.2)$$

Параметр a_1 – коэффициент при x – именуется коэффициентом регрессии. Он показывает, на какую величину в абсолютном выражении изменится результативный показатель y при изменении факторного показателя x на единицу.

Зная a_1 , легко определить второй параметр уравнения регрессии как $a_0 = \bar{y} - a_1 \bar{x}$.

Если связь выражена параболой второго порядка

$$\bar{y}_x = a_0 + a_1 x + a_2 x^2,$$

то система нормальных уравнений для отыскания параметров a_0 , a_1 , a_2 выглядит следующим образом:

$$\begin{cases} na_0 + a_1 \sum x + a_2 \sum x^2 = \sum y; \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 = \sum xy; \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 = \sum x^2 y. \end{cases} \quad (4.3)$$

Нахождение уравнений регрессии, как правило, сопровождается измерением тесноты связи (зависимости).

Измерить **тесноту корреляционной зависимости** – значит определить, насколько вариация результативного показателя вызвана вариацией факторного признака [4, 6, 10, 20]. Эта задача для всех форм связи может быть решена с помощью исчисления теоретического корреляционного отношения (η):

$$\eta = \frac{\delta}{\sigma_y} = \sqrt{\frac{\delta^2}{\sigma_y^2}}, \quad (4.4)$$

где $\delta^2 = \frac{\sum (\bar{y}_x - \bar{y})^2}{n}$ – дисперсия в ряду выровненных значений результативного показателя \bar{y}_x ;

$$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n} - \text{дисперсия в ряду фактических значений } y.$$

Так как дисперсия δ^2 отражает вариацию в ряду \bar{y}_x только за счет вариации фактора x , а дисперсия σ_y^2 отражает вариацию y за счет всех факторов, то их отношение, именуемое теоретическим коэффициентом детерминации, показывает, какой удельный вес в общей дисперсии ряда y занимает дисперсия, вызываемая вариацией фактора x . Квадратный корень из отношения этих дисперсий дает нам теоретическое корреляционное отношение

$$\eta = \sqrt{\frac{\delta^2}{\sigma_y^2}}.$$

Если $\delta^2 = \sigma_y^2$, то это означает, что роль других факторов в вариации y сведена на нет, и отношение $\eta = \frac{\delta}{\sigma_y} = 1$ означает полную зависимость вариации x . Если $\delta^2 = 0$, то это значит, что вариация x никак не влияет на вариацию y , и в этом случае $\eta = \frac{\delta}{\sigma_y} = 0$. Следовательно, максимальное значение, которое может принимать корреляционное отношение, равно 1, минимальное значение – 0.

Дисперсию теоретических значений результативного показателя (т.е. δ^2) часто называют *факторной*, поскольку она отражает влияние вариации фактора x на вариацию y и обозначается как δ_{ϕ}^2 или $D_{\bar{y}_x}$.

Согласно правилу сложения дисперсий общую дисперсию эмпирических значений результативного показателя y (σ_y^2) можно представить как сумму *факторной дисперсии* (δ_{ϕ}^2) и так называемой *остаточной дисперсии* ($\sigma_{ост}^2$), отражающей вариацию результативного показателя за счет всех остальных факторов (кроме x), не учтенных в уравнении регрессии [2, 6, 10], т.е.

$$\sigma_y^2 = \delta_{\phi}^2 + \sigma_{ост}^2.$$

Выразив факторную дисперсию, как $\delta_{\phi}^2 = \sigma_y^2 - \sigma_{ост}^2$ и подставив это выражение в формулу, получим еще одну формулу для вычисления корреляционного отношения:

$$\eta = \sqrt{\frac{\sigma_y^2 - \sigma_{ост}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}}. \quad (4.5)$$

В данном виде корреляционное отношение при криволинейной зависимости обычно называют **индексом корреляции**.

В формуле (4.5) используются дисперсия эмпирического ряда игроков и остаточная дисперсия, рассчитываемая по формуле

$$\sigma_{ост}^2 = \frac{\sum (y_i - \bar{y}_x)^2}{n}. \quad (4.6)$$

Математически легко доказывается, что в случае линейной зависимости корреляционное отношение $\eta = \frac{\delta_{\phi}}{\sigma_y}$ может быть преобразовано в выражение

$a_1 \frac{\sigma_x}{\sigma_y}$, которое называют **линейным коэффициентом корреляции** и

обозначают r , т.е. $r = a_1 \frac{\sigma_x}{\sigma_y}$, где a_1 – коэффициент регрессии в уравнении

связи, σ_x и σ_y – среднее квадратичное отклонение соответственно в ряду x и в ряду y .

Линейный коэффициент корреляции можно выразить и другими формулами, тождественными первой, в частности [12–14]:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}; \quad r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{[\bar{x}^2 - (\bar{x})^2][\bar{y}^2 - (\bar{y})^2]}}; \quad r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}$$

или $r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}},$

а также
$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}}$$
.

Линейный коэффициент корреляции может принимать по модулю значения от 0 до 1 (знак «+» при прямой зависимости и знак «-» при обратной зависимости).

Рассмотрим решение некоторых примеров по этой теме.

Пример 1

Пусть по 10 однотипным предприятиям имеются следующие данные о выпуске продукции (x) в тысячах единиц и о расходе условного топлива (y) в тоннах (графы 1 и 2 табл. 4.1).

Таблица 4.1

Данные о выпуске продукции

x	y	x^2	xy	$\bar{y}_x = 1,16 + 0,547x$	y^2
1	2	3	4	5	6
5	4	25	20	3,9	16
6	4	36	24	4,4	16
8	6	64	48	5,5	36
8	5	64	40	5,5	25
10	7	100	70	6,6	49
10	8	100	80	6,6	64
14	8	196	112	8,8	64
20	10	400	200	12,1	100
20	12	400	240	12,1	144
24	16	576	384	14,3	256
125	80	1961	1218	80	770

Требуется найти уравнение зависимости расхода топлива от выпуска продукции (или уравнение регрессии y по x) и измерить тесноту зависимости между ними.

Решение.

1. Рассматривая уравнение регрессии в форме линейной функции вида $y_x = a_0 + a_1x$, параметры данного уравнения (a_0 и a_1) найдем из системы нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum x = \sum y, \\ a_0 \sum x + a_1 \sum x^2 = \sum xy. \end{cases}$$

Необходимые для решения суммы $\sum x, \sum y, \sum x^2, \sum xy$ рассчитаны выше в табл. 4.1. Подставляем их в уравнения и решаем систему:

$$\begin{cases} 10a_0 + 125a_1 = 80, \\ 125a_0 + 1961a_1 = 1218. \end{cases}$$

$$a_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{10 \cdot 1218 - 125 \cdot 80}{10 \cdot 1961 - 125^2} = \frac{2180}{3985} = 0,547.$$

$$a_0 = \bar{y} - a_1 \bar{x} = \frac{80}{10} - 0,547 \cdot \frac{125}{10} = 8 - 6,84 = 1,16.$$

Отсюда $y_x = 1,16 + 0,547x$.

Подставляя в это уравнение последовательно значения $x = 5, 6, 8, 10$ и т. д., получаем выровненные (теоретические) значения результативного показателя y_x (графа 5 табл. 4.1).

Поскольку параметры уравнения регрессии являются оценочными, то для каждого из них рассчитывается средняя ошибка, т.е. μ_{a_i} .

2. Для измерения тесноты зависимости между y и x воспользуемся, прежде всего, линейным коэффициентом корреляции (поскольку зависимость рассматривалась линейной):

а) применяем формулу
$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}.$$

Находим $\bar{x}\bar{y} = 121,8$; $\bar{x} = 12,5$; $\bar{y} = 8$; $\bar{x}^2 = 196,1$.

Определяем σ_x и σ_y , предварительно найдя $\sum y^2 = 770$ и $\bar{y}^2 = 77$.

$$\sigma_x = \sqrt{\bar{x}^2 - (\bar{x})^2} = \sqrt{196,1 - 12,5^2} = \sqrt{39,85} = 6,31;$$

$$\sigma_y = \sqrt{\bar{y}^2 - (\bar{y})^2} = \sqrt{77 - 8^2} = \sqrt{13} = 3,6.$$

Отсюда $r = \frac{121,8 - 12,5 \cdot 8}{6,31 \cdot 3,6} = 0,96$.

Значение линейного коэффициента корреляции $r = 0,96$ (т.е. близкое к единице) характеризует не только меру тесноты зависимости вариации y от вариации x , но и степень близости этой зависимости к линейной;

б) воспользуемся еще одной формулой линейного коэффициента корреляции:

$$r = a_1 \frac{\sigma_x}{\sigma_y} = 0,547 \cdot \frac{6,31}{3,6} = 0,96,$$

т.е. результат тот же.

При расчете коэффициента корреляции очень важно оценить его значимость. Оценка значимости (существенности) линейного коэффициента корреляции основана на сопоставлении значения r с его средней квадратичной ошибкой.

Средняя ошибка коэффициента корреляции при $n > 50$ рассчитывается приближенно по формуле

$$\sigma_r = \frac{1 - r^2}{\sqrt{n}}.$$

Если при этом коэффициент корреляции r превышает свою среднюю ошибку σ_r больше чем в 3 раза, т.е. если $\frac{|r|}{\sigma_r}$, то он считается значимым; а связь – реальной.

При $n < 30$ значимость коэффициента корреляции проверяется на основе t -критерия Стьюдента. Для этого рассчитывается фактическое (расчетное) значение критерия:

$$t_{\text{факт}} = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}},$$

которое сопоставляется с $t_{\text{табл}}$, определяемым по прил. 9 для числа степеней свободы $\nu = n - 2$ и заданного уровня значимости (обычно $\alpha = 0,05$).

Если $t_{\text{факт}} > t_{\text{табл}}$, коэффициент корреляции r считается значимым, а связь – реальной. Если $t_{\text{факт}} < t_{\text{табл}}$, то считается, что связь между x и y отсутствует и значение r , отличное от нуля, получено случайно.

В рассматриваемом примере средняя ошибка коэффициента корреляции

$$\sigma_r = \frac{\sqrt{1-r^2}}{n-2} = \frac{\sqrt{1-0,96^2}}{10-2} = \frac{\sqrt{0,0784}}{2,83} = 0,1,$$

$$\text{а } t_{\text{факт}} = \frac{r}{\sigma_r} = \frac{0,96}{0,1} = 9,6.$$

По таблице прил. 9 находим, что при числе степеней свободы $\nu = 10 - 2 = 8$ и уровне значимости $\alpha = 0,05$ табличное (критическое, пороговое) t равно 2,306, т.е. $t_{\text{табл}} = 2,306$.

Поскольку фактическое (расчетное) t больше табличного, т.е. $t_{\text{факт}} > t_{\text{табл}}$, то линейный коэффициент корреляции $r = 0,96$ считается значимым, а связь между x и y – реальной.

3. Для измерения тесноты зависимости воспользуемся теоретическим корреляционным отношением:

$$\eta = \sqrt{\frac{\delta_\phi^2}{\sigma_y^2}} = \sqrt{\frac{D_{\bar{y}_x}}{D_y}},$$

где δ_ϕ^2 и σ_y^2 – дисперсии соответственно теоретических и эмпирических значений результативного показателя.

Расчет их показан ниже в табл. 4.2.

Таблица 4.2

Результаты расчета

x	y	\bar{y}_x	$y - \bar{y}$	$(y - \bar{y})^2$	$\bar{y}_x - \bar{y}$	$(\bar{y}_x - \bar{y})$	$y - \bar{y}_x$	$(y - \bar{y}_x)^2$
1	2	3	4	5	6	7	8	9
5	4	3.90	-4	16	-4.10	16.8100	0.10	0.100
6	4	4.44	-4	16	-3.56	12.6736	-0.44	0.1936
8	6	5.54	-2	4	-2.46	6.0516	0.46	0.2116
8	5	5.54	-3	9	-2.46	6.0516	-0.54	0.2916
10	7	6.63	-1	1	-1.37	1.8769	0.37	0.1369
10	7	6.63	0	0	-1.37	1.8769	1.37	1.8769
14	8	8.82	0	0	0.82	0.6724	-0.82	0.6724
20	10	12.10	2	40	4.10	16.8100	-2.10	4.4100
20	12	12.10	4	16	4.10	16.8100	-0.10	0.0100
24	16	14.30	8	64	6.30	39.6900	1.70	2.8900
125	80	80	0	130	0	119,3230	0	10,7030

Дисперсия выровненных значений результативного показателя, или факторная дисперсия, $\delta_{\phi}^2 = \frac{\sum (\bar{y}_x - \bar{y})^2}{n} = \frac{119,323}{10} = 11,932$; общая дисперсия эмпирических значений результативного показателя

$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n} = \frac{130}{10} = 13$; теоретическое корреляционное отношение

$$\eta = \sqrt{\frac{\delta_{\phi}^2}{\sigma_y^2}} = \sqrt{\frac{11,932}{13}} = 0,958 \approx 0,96.$$

Рассчитанные показатели позволяют сделать вывод о том, что связь между вариацией результативного показателя (y) и факторного (x) весьма большая.

Вместо дисперсии выровненных значений y , т.е. δ_ϕ^2 , можно воспользоваться **остаточной дисперсией**:

$$\sigma_{ост}^2 = \frac{\sum (y - \bar{y}_x)^2}{n} = \frac{10,703}{10} = 1,07.$$

(Расчет $\sum (y - \bar{y}_x)$ показан в графах 8 и 9 табл. 4.2)

Искомое корреляционное отношение (или индекс корреляции)

$$\eta = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}} = \sqrt{1 - \frac{1,07}{13}} = 0,96.$$

Остаточная дисперсия, вернее, корень квадратный из нее, т.е. $\sigma_{ост}$, используется для расчета *средних ошибок параметров* уравнения регрессии.

Так, средняя ошибка параметра a_0 составляет величину $\mu_{a_0} = \frac{\sigma_{ост}}{\sqrt{n-2}}$, а для a_1

$$\mu_{a_1} = \frac{\sigma_{ост}}{\sigma_x \sqrt{n-2}}.$$

Сопоставляя значение параметра с его средней ошибкой по значению

$t = \frac{a_i}{\mu_{a_i}}$, судят о значимости данного параметра.

Если число наблюдений $n > 20$, то параметр считается значимым при $t > 3$.

Если $n < 20$, то обращаются к специальным таблицам значений t -критерия Стьюдента (см. прил. 9). И в данном случае параметр считается значимым, если $t_{факт} > t_{табл}$.

В рассмотренном примере для уравнения регрессии $\bar{y}_x = 1,16 + 0,547x$ ошибки параметров:

$$\mu_{a_0} = \frac{\sigma_{ост}}{\sqrt{n-2}} = 0,36, \text{ т.е. } a_0 = 1,16 \pm 0,36.$$

$$\mu_{a_1} = \frac{\sigma_{ост}}{\sigma_x \sqrt{n-2}} = 0,058, \text{ т.е. } a_1 = 0,547 \pm 0,058.$$

Чтобы сделать вывод о значимости параметров, находим

$$t_{a_0} = \frac{1,16}{0,36} = 3,2 \text{ и } t_{a_1} = \frac{0,547}{0,058} = 9,43.$$

Так как $n < 20$, обращаемся к таблице значений t -критерия (см. прил. 9). Для $\alpha = 0,05$ и $\nu = 10 - 2 = 8$ находим $t_{\text{табл}} = 2,306$. Поскольку $t_{\text{факт}} > 2,306$, то считаем параметры значимыми.

В приведенных формулах ошибок параметров μ_{a_0} и μ_{a_1} вместо остаточной дисперсии $\sigma_{\text{ост}}^2$ можно пользоваться значением линейного коэффициента корреляции (или корреляционного отношения η), исходя из следующих преобразований:

$$\sigma_{\text{ост}}^2 = \sigma_y^2 - \delta_\phi^2;$$

разделив обе части равенства на σ_y^2 , получим $\frac{\sigma_{\text{ост}}^2}{\sigma_y^2} = 1 - r^2$. Отсюда

$$\sigma_{\text{ост}} = \sigma_y \sqrt{1 - r^2}. \text{ Тогда } \mu_{a_1} = \frac{\sigma_y \sqrt{1 - r^2}}{\sqrt{n - 2}}, \text{ а } \mu_{a_0} = \frac{\sigma_y \sqrt{1 - r^2}}{\sigma_x \sqrt{n - 2}}.$$

Наряду с проверкой значимости отдельных параметров, можно оценить значимость уравнения регрессии в целом.

Эта задача решается на основе расчета F -критерия Фишера и сопоставления его с табличным (критическим), F -критерий представляет собой отношение факторной дисперсии (δ_ϕ^2) к остаточной ($\delta_{\text{ост}}^2$), каждая из которых рассчитана на одну степень свободы:

$$F = \frac{\delta_\phi^2 / (m - n)}{\sigma_{\text{ост}}^2 / (n - m)} = \frac{\delta_\phi^2}{\sigma_{\text{ост}}^2} \cdot \frac{n - m}{m - 1},$$

где m – число параметров в уравнении регрессии;

$(m - 1)$ – число степеней свободы для факторной дисперсии;

n – число наблюдений;

$(n - m)$ – число степеней свободы для остаточной дисперсии.

Часто формула F-критерия записывается в следующем виде:

$$F = \frac{\frac{\delta_{\phi}^2}{k}}{\frac{\sigma_{ост}^2}{(n-k-1)}} = \frac{\delta_{\phi}^2}{\sigma_{ост}^2} \cdot \frac{n-k-1}{k},$$

где k – число коэффициентов регрессии, которое на единицу меньше числа параметров, т.е. $k = m - 1$.

F-критерий можно рассчитать и через коэффициент корреляции (r или η):

$$F = \frac{r^2}{1-r^2} \cdot \frac{n-m}{m-1} \quad (\text{или } F = \frac{r^2}{1-r^2} \cdot \frac{n-k-1}{k}).$$

Расчетное F сопоставляется с табличным, определяемым по прил. 8 для числа степеней свободы $\nu_1 = m - 1$ и $\nu_2 = n - m$ при заданном уровне значимости (например, $\alpha = 0,05$).

Если уравнение считается значимым, для рассмотренного выше примера

$$\bar{y}_x = 1,16 + 0,547x;$$

$$F_{расч} = \frac{\delta_{\phi}^2}{\sigma_{ост}^2} \cdot \frac{n-m}{m-1} = \frac{11,932}{1,07} \cdot \frac{10-2}{2-1} = 89,2.$$

По прил. 8 находим $F_{табл}$; при $\alpha = 0,05$ и $\nu_1 = 2 - 1 = 1$, $\nu_2 = 10 - 2 = 8$ имеем $F_{табл} = 5,32$.

Так как $F_{расч} > F_{табл}$, то уравнение значимо.

Контрольные вопросы

1. В чем сущность корреляционной связи между показателями?
2. Как определяются параметры уравнения регрессии при линейной зависимости?
3. Как определяются ошибки параметров уравнения регрессии?
4. Как проверяется значимость уравнения регрессии?
5. Какие вы знаете показатели измерений тесноты зависимости?
6. Как оценивается значимость коэффициента корреляции, рассчитанного по выборочным данным?

5. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

5.1. Понятие статистической гипотезы

Часто бывает необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но имеются основания предположить, что он имеет определенный вид (назовем его A), выдвигают гипотезу: генеральная совокупность распределена по закону A . Таким образом, в этой гипотезе речь идет о виде предполагаемого распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр Φ равен определенному значению Φ_0 , выдвигают гипотезу: $\Phi = \Phi_0$. Таким образом, в этой гипотезе речь идет о предполагаемой величине параметра одного известного распределения. Возможны и другие гипотезы: о равенстве параметров двух или нескольких распределений, о независимости выборок и многие другие.

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Например, статистическими являются гипотезы:

- генеральная совокупность распределена по закону Пуассона;
- дисперсии двух нормальных совокупностей равны между собой.

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй – о параметрах двух известных распределений. Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то имеет место противоречащая гипотеза. По этой причине эти гипотезы целесообразно различать.

Нулевой (основной) называют выдвинутую гипотезу H_0 . Конкурирующей (альтернативной) называют гипотезу H_1 , которая противоречит нулевой. Например, если нулевая гипотеза состоит в предположении, что математическое ожидание нормального распределения $a = 10$, то конкурирующая гипотеза, в частности, может состоять в предположении, что $a \neq 10$. Коротко это записывают так:

$$H_0: a = 10; H_1 \neq 10.$$

Различают гипотезы, которые содержат только одно или более одного предположений. Простой называют гипотезу, содержащую только одно предположение. Например, если λ – параметр показательного распределения, то гипотеза $\lambda = 5$ – простая. Гипотеза H_0 : математическое ожидание нормального распределения равно 3 (σ известно) – простая.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез. Например, сложная гипотеза $H: \lambda > 5$ состоит из бесчисленного множества простых гипотез вида $H_i: \lambda = b_i$ где b_i – любое число, большее 5. Гипотеза H_0 : математическое ожидание нормального распределения равно 3 (σ неизвестно) – сложная.

5.2. Теоретическая схема проверки гипотез

Рассмотрим случай, когда некоторая оценка $\hat{\Phi}$ вычислена по выборке объема n независимых наблюдений случайной величины x . Предположим, что есть основания считать истинное значение оцениваемого параметра Φ равным некоторой величине Φ_0 . Даже если $\Phi = \Phi_0$, то выборочное значение $\hat{\Phi}$ не будет, по-видимому, точно совпадать с Φ_0 из-за выборочной изменчивости статистики $\hat{\Phi}$. Поэтому возникает следующий вопрос. Если принять гипотезу $\Phi = \Phi_0$, то насколько велико должно быть различие между $\hat{\Phi}$ и Φ_0 , чтобы эту гипотезу следовало отвергнуть как ошибочную? На этот вопрос можно ответить в статистическом смысле, рассматривая вероятность достижения некоторой заданной разности между $\hat{\Phi}$ и Φ_0 на основе выборочного распределения параметра $\hat{\Phi}$. Если вероятность превышения этой разности $\hat{\Phi}$ и заданного уровня Φ_0 мала, то этот уровень можно считать значимым и гипотезу $\Phi = \Phi_0$ следует отвергнуть. Если вероятность превышения данной разности не является малой, то наличие этой разности можно объяснить обычной статистической изменчивостью и гипотезу $\Phi = \Phi_0$ считать правдоподобной. Природа статистических выводов такова, что при отклонении гипотезы мы

можем заранее оценить вероятность возможной ошибки (отклонить истинную гипотезу). Напротив, если гипотеза не отклонена, то это не означает, что она подтверждена с заданной вероятностью. Это означает лишь, что она согласуется с поставленным статистическим опытом, но, возможно, осуществим другой опыт, в результате которого она должна быть отвергнута.

Приведенные рассуждения намечают самую простую форму статистического метода, называемого проверкой гипотез. Для того чтобы пояснить сущность этого метода, предположим, что выборочная величина $\hat{\Phi}$, представляющая собой несмещенную оценку параметра Φ , имеет плотность распределения $p(\Phi)$. Если гипотеза $\Phi = \Phi_0$ верна, то функция $p(\hat{\Phi})$ должна обладать средним значением Φ_0 , как показано на рис. 5.2.1. Вероятность того, что параметр Φ не будет превышать нижнего уровня $\Phi_1 - a/2$, составит

$$P[\hat{\Phi} \leq \Phi_{1-a/2}] = \int_{-\infty}^{\Phi_{1-a/2}} p(\hat{\Phi}) d\hat{\Phi} = \frac{a}{2}. \quad (5.1)$$

Вероятность того, что параметр Φ превысит верхний уровень $\Phi_{a/2}$, равна

$$P[\hat{\Phi} > \Phi_{a/2}] = \int_{\Phi_{a/2}}^{\infty} p(\hat{\Phi}) d\hat{\Phi} = \frac{a}{2}. \quad (5.2)$$

Следовательно, вероятность того, что параметр $\hat{\Phi}$ выйдет за пределы интервала с границами $\Phi_{1-a/2}$ и $\Phi_{a/2}$, составляет α . Теперь примем величину α настолько малой, чтобы попадание параметра $\hat{\Phi}$ за пределы интервала с границами $\Phi_{1-a/2}$ и $\Phi_{a/2}$ было маловероятным. Если после извлечения выборки и определения величины $\hat{\Phi}$ окажется, что эта величина выходит за пределы интервала $(\Phi_{1-a/2}; \Phi_{a/2})$, то в этом случае есть серьезные основания подвергнуть сомнению справедливость проверяемой гипотезы $\Phi = \Phi_0$. Действительно, если гипотеза верна, такое значение Φ будет маловероятным. Следовательно, гипотезу равенства величин Φ и Φ_0 следует отвергнуть. С другой стороны, если параметр $\hat{\Phi}$ попадает в интервал $(\Phi_{1-a/2}; \Phi_{a/2})$, то в этом случае нет серьезных оснований подвергать сомнению справедливость

проверяемой гипотезы. Следовательно, гипотезу равенства $\Phi = \Phi_0$ можно принять.

Малое значение вероятности α , используемое при проверке гипотезы, называют уровнем значимости критерия. Интервал значений $\hat{\Phi}$, при которых гипотезу следует отвергнуть, называют областью отклонения гипотезы, или критической областью. Интервал значений $\hat{\Phi}$, при которых гипотезу можно принять, носит название области принятия гипотезы.

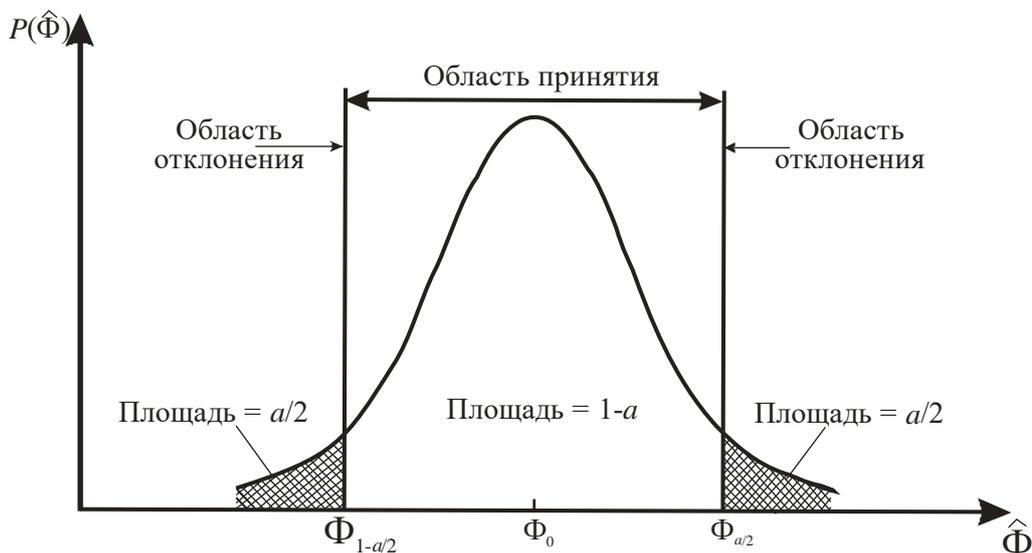


Рис. 5.1. Области принятия и отклонения при проверке гипотез:

$\hat{\Phi}$ – оценка параметра Φ ;

$p(\hat{\Phi})$ – плотность распределения оценки параметра Φ ;

Φ_0 – истинное значение параметра Φ ;

$\Phi_{1-\alpha/2}$ – нижний уровень оценки;

$\Phi_{\alpha/2}$ – верхний уровень оценки;

α – вероятность выхода оценки за пределы верхнего и нижнего уровней

Описанный простой прием проверки гипотез называется двусторонним критерием, так как если гипотеза не верна, величина Φ может быть как больше, так и меньше Φ_0 . Поэтому необходимо проверить значимость расхождения между Φ и Φ_0 с обеих сторон. В других случаях может оказаться достаточным односторонний критерий.

Рассмотрим, например, гипотезу, заключающуюся в том, что $\Phi \geq \Phi_0$. В этом случае гипотеза может оказаться ошибочной только при Φ , меньшем чем

Φ_0 . Поэтому критерий следует изменить, используя нижнюю границу плотности распределения $p(\Phi)$.

При проверке гипотез возможны ошибки двух родов. Во-первых, гипотеза отвергается, когда в действительности она верна. Эта возможная ошибка называется ошибкой первого рода. Во-вторых, гипотеза принимается, когда она не верна. Эта возможная ошибка называется ошибкой второго рода. Как видно из рис. 5.1, ошибка первого рода допускается, если гипотеза верна, а параметр Φ попадает в область отклонения гипотезы. Отсюда следует, что вероятность допустить ошибку первого рода равна α , т.е. уровню значимости критерия.

Для того чтобы найти, какова вероятность допустить ошибку второго рода, необходимо задать определенную величину отклонения истинного значения Φ от гипотетического значения параметра Φ_0 , которое требуется определить. Предположим, например, что истинное значение параметра Φ в действительности равно $\Phi_0 + d$ или $\Phi_0 - d$, как показано на рис. 5.2. Если, согласно гипотезе, $\Phi = \Phi_0$, а в действительности $\Phi = \Phi_0 \pm d$, то вероятность того, что Φ попадает в область принятия гипотезы, т.е. в интервал

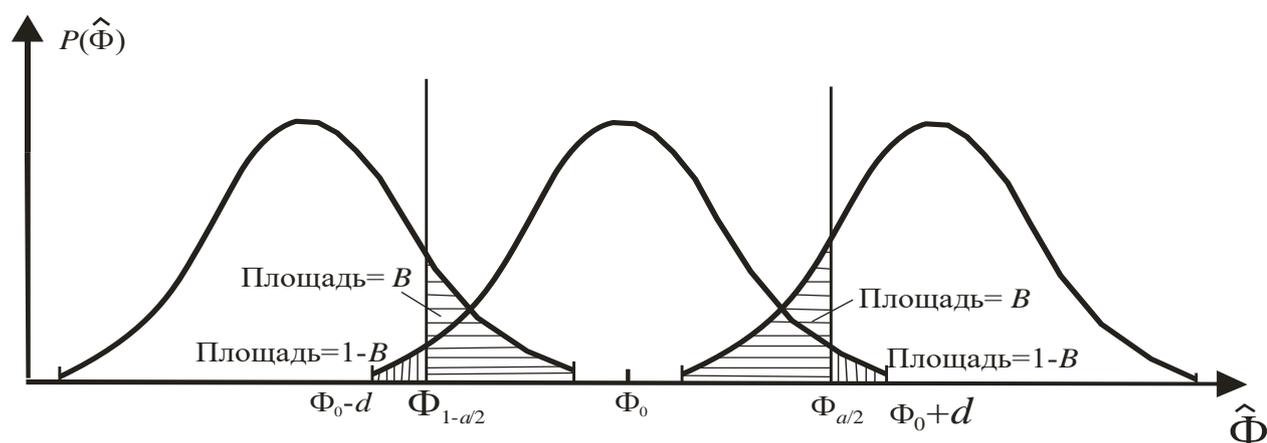


Рис. 5.2. Области принятия и отклонения, соответствующие ошибке второго рода при проверке гипотезы: d – смещение истинного значения параметра Φ_0 ;
 β – вероятность попадания в область принятия гипотезы при смещении истинного значения; $1 - \beta$ – мощность критерия

$(\Phi_{1-\alpha/2}; \Phi_{\alpha/2})$, составляет β . Это значит, что вероятность допустить ошибку второго рода при выявлении отклонения $\pm d$ от гипотетического значения Φ_0 равна β . Вероятность $1 - \beta$ носит название мощности критерия. Очевидно, что при любом заданном объеме выборки N вероятность допустить ошибку первого рода можно уменьшить, уменьшая уровень значимости α . Однако при этом увеличивается вероятность β допустить ошибку второго рода (снижается мощность критерия). Единственный способ уменьшить и α , и β состоит в том, чтобы увеличить объем выборки N , по которой находят оценку ϕ . Исходя из этих соображений и определяется объем выборок, необходимый для проведения статистических исследований [4, 7, 10, 20].

Пример. Предположим, что есть основания считать среднее значение μ_x случайной величины x равным 10. Пусть далее известна дисперсия величины x : $\sigma_x^2 = 4$. Найдем, каков должен быть объем выборки для проверки гипотезы $\mu_x = 10$ при 5%-м уровне значимости, причем вероятность допустить ошибку второго рода при определении 10%-го отклонения от гипотетической величины также должна составить 5%. Определим при этих условиях область принятия, которую следует использовать при проверке гипотезы.

Решение. Несмещенная оценка величины μ_x равна выборочному среднему \bar{x} . Соответствующее выборочное распределение параметра \bar{X}

$$\bar{X} = \frac{\sigma_x}{\sqrt{N}} Z + \mu_x,$$

где z – нормально распределенная величина с нулевым средним и единичной дисперсией. Заметим, что это выборочное распределение величины x является точным, если x имеет нормальное распределение, и представляет собой хорошее приближение, если распределение величины x отличается от нормального.

Верхняя и нижняя границы области принятия гипотезы в рассматриваемом случае имеют вид:

$$\text{верхняя граница} = \frac{\sigma_x}{\sqrt{N}} Z_{\alpha/2} + \mu_x;$$

$$\text{нижняя граница} = \frac{\sigma_x}{\sqrt{N}} Z_{1-\alpha/2} + \mu_x.$$

Пусть теперь истинное среднее значение $\mu_x = \mu_x \pm d$. Тогда если выборочное значение \bar{X} лежит ниже верхней границы или выше нижней границы, то ошибка второго рода допускается с вероятностью β . Используя выборочные распределения величин $\mu_x = \mu_x + d$ и $\mu_x = \mu_x - d$, можно найти, что

$$\text{верхняя граница} = \frac{\sigma_x}{\sqrt{N}} z_{1-\beta} + \mu_x + d;$$

$$\text{нижняя граница} = \frac{\sigma_x}{\sqrt{N}} z_{\beta} + \mu_x - d.$$

Следовательно, справедливы равенства

$$\frac{\sigma_x}{\sqrt{N}} z_{\alpha/2} + \mu_x = \frac{\sigma_x}{\sqrt{N}} z_{1-\beta} + \mu_x + d;$$

$$\frac{\sigma_x}{\sqrt{N}} z_{1-\alpha/2} + \mu_x = \frac{\sigma_x}{\sqrt{N}} z_{\beta} + \mu_x - d.$$

Эти соотношения можно объединить следующим образом:

$$z_{\alpha/2} = z_{1-\beta} + \frac{\sqrt{N}}{\sigma_x} d = -z_{\beta} + \frac{\sqrt{N}}{\sigma_x} d.$$

Отсюда следует, что искомый объем выборки

$$N = \left[\frac{\sigma_x (z_{\alpha/2} + z_{\beta})}{d} \right]^{-2}.$$

Для условий настоящего примера $\sigma_x = 2$; $z_{\alpha/2} = 1,96$; $z_{\beta} = 1,465$; $\beta = 0,1 \cdot 10 = 1$ находим, что искомый объем выборки $N = 52$. Область принятия гипотезы в рассматриваемом случае имеет следующие границы:

$$\text{верхняя граница} = \frac{\sigma_x}{\sqrt{N}} z_{\alpha/2} + \mu_x \approx 10,54;$$

$$\text{нижняя граница} = \frac{\sigma_x}{\sqrt{N}} z_{1-\alpha/2} + \mu_x \approx 9,46.$$

5.3. Проверка типовых статистических гипотез

Проверка гипотезы о равенстве математического ожидания заданному значению. Пусть сформулированы нулевая гипотеза $H_0: M(y) = \bar{Y}$ и альтернативная гипотеза $H_1: M(y) \neq \bar{Y}$. Выборка наблюдений объемом N для проверки нулевой гипотезы осуществляется из генеральной совокупности

значений случайной величины y , распределенных по нормальному закону. Оценкой математического ожидания по выборке будет среднее арифметическое \bar{y} :

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i,$$

а оценкой дисперсии – s^2

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2.$$

Для проверки нулевой гипотезы в данном случае рекомендуется проверка по критерию Стьюдента. Наблюдаемое значение критерия Стьюдента вычисляется по следующей формуле:

$$t_{\text{набл}} = \frac{|\bar{y} - \bar{Y}|}{S\sqrt{N}}.$$

Критическое значение критерия $t_{\text{кр}}$ определяется для выбранной доверительной вероятности $P = 1 - \alpha$ и заданного объема выборки N из таблиц критических значений критерия Стьюдента. При $t_{\text{набл}} < t_{\text{кр}}$ можно считать, что данные выборки не противоречат нулевой гипотезе; если $t_{\text{набл}} > t_{\text{кр}}$, то гипотеза отвергается.

Проверка гипотезы о равенстве средних значений. Пусть выборки наблюдений объемом N_1 и N_2 берутся из двух генеральных совокупностей с нормальным распределением, причем дисперсии генеральных совокупностей σ_1 и σ_2 равны. Необходимо проверить гипотезу H_0 о равенстве математических ожиданий $M(Y_1)$ и $M(Y_2)$.

По данным выборок определяются оценки математических ожиданий и дисперсий $\bar{y}_1, \bar{y}_2, s_1^2, s_2^2$. Объединенная оценка дисперсии генеральных совокупностей

$$s^2 = [s_1^2(N_1 - 1) + s_2^2(N_2 - 1)] / (N_1 + N_2 - 1). \quad (5.3)$$

Для проверки нулевой гипотезы вычисляется наблюдаемое значение критерия Стьюдента:

$$t_{\text{набл}} = \frac{|\bar{y}_1 - \bar{y}_2|}{s} \cdot \sqrt{\frac{N_1 N_2}{N_1 + N_2}}. \quad (5.4)$$

Критическое значение критерия определяется для данной доверительной вероятности P и $N = N_1 + N_2$. При $t_{\text{набл}} < t_{\text{кр}}$ гипотеза принимается; если $t_{\text{набл}} > t_{\text{кр}}$, то гипотеза отвергается.

Пример. В одинаковых условиях было обработано по 25 втулок развертками диаметром 6 и 10 мм. Результаты измерений показали, что средняя величина разбивки отверстий (разность между диаметром развертки и диаметром отверстия) для разверток диаметром 6 мм составляет 10,9 мкм, а диаметром 10 мм – 9,8 мкм. Оценки дисперсий в обоих случаях составляют соответственно $s_1^2 = 3,8$ мкм²; $s_2^2 = 4,76$ мкм². Необходимо установить, влияет ли диаметр развертки на разбивку отверстия.

Решение. Согласно формулам (5.3) и (5.4) определяем

$$s^2 = \frac{3,8 \cdot 24 + 4,76 \cdot 24}{49} = 4,19; \quad t_{\text{набл}} = \frac{10,9 - 9,8}{\sqrt{4,19}} \cdot \sqrt{\frac{25 \cdot 25}{25 + 25}} = 1,9.$$

По таблицам критерия Стьюдента находим при $P = 0,95$ и $N = 50$: $t_{\text{кр}} = 2,01$. Поскольку $t_{\text{набл}} = 1,9 < t_{\text{кр}} = 2,01$, можно считать, что изменение диаметра развертки в пределах от 6 до 10 мм не оказывает существенного влияния на величину разбивки развернутого отверстия.

Проверка гипотезы о равенстве дисперсий. При обработке экспериментальных данных часто требуется выяснить вопрос об однородности выборочных дисперсий, т.е. об их равенстве дисперсии генеральной совокупности.

Пусть для двух независимых выборок из нормальной генеральной совокупности с объемами N_1 и N_2 вычислены оценки s_1^2 и s_2^2 . Требуется проверить нулевую гипотезу $H_0: \sigma_1^2 = \sigma_2^2$ относительно альтернативной гипотезы $H_1: \sigma_1^2 > \sigma_2^2$. Проверка проводится при помощи критерия Фишера F .

Наблюдаемое значение критерия $F_{\text{набл}} = s_1^2/s_2^2$, ($s_1^2 > s_2^2$) сравнивается с критическим по таблицам критических значений критерия Фишера для

выбранных доверительной вероятности и объемов выборок N_1 и N_2 . Если $F_{\text{набл}} < F_{\text{кр}}$, то выборочные данные не противоречат нулевой гипотезе.

При анализе выборочных данных могут выдвигаться гипотезы об однородности дисперсий в нескольких выборках. В этом случае можно использовать критерий Кохрена. Наблюдаемое значение критерия $G_{\text{набл}}$ определяется по формуле

$$G_{\text{набл}} = S_{\text{imax}}^2 / \sum_{i=1}^n S_i^2, \quad (5.5)$$

где S_{imax}^2 – максимальная оценка дисперсии среди n сравниваемых дисперсий (все n выборок имеют одинаковый объем N).

Критическое значение критерия определяется из таблицы критических значений критерия Кохрена в зависимости от принятых доверительной вероятности P , объема выборок m и их числа n .

Пример 1. Проверить гипотезу о равенстве дисперсий в предыдущем примере о проверке отверстий во втулках.

Решение. Определяем наблюдаемое значение критерия Фишера:

$F_{\text{набл}} = 4,76/3,8 = 1,25$. Из таблицы для $P = 0,95$, $N_1 = N_2 = 20$ находим $F_{\text{кр}} = 2,16$, а для $N_1 = N_2 = 30$ значение $F_{\text{кр}} = 1,84$. Искомое значение $F_{\text{кр}}$ для $N_1 = N_2 = 25$ находим линейной интерполяцией: $F_{\text{кр}} = (1,84 + 2,16)/2 = 2$. Так как $F_{\text{набл}} < F_{\text{кр}}$ ($1,25 < 2$), то нулевая гипотеза о равенстве дисперсий принимается.

Пример 2. С четырех автоматов, настроенных на обработку одних и тех же деталей, взято по одной текущей выборке объемом $N = 10$. Оценки дисперсий их размеров имели следующие значения: $s_1^2 = 106$ мкм²; $s_2^2 = 294$ мкм²; $s_3^2 = 216$ мкм²; $s_4^2 = 410$ мкм².

Требуется определить, одинакова ли точность автоматов, т.е. можно ли принять гипотезу об однородности дисперсий.

Решение. Согласно формуле (5.5)

$$G_{\text{набл}} = \frac{410}{106 + 294 + 216 + 410} = 0,3996.$$

По таблице критических значений критерия Кохрена для $P = 0,95$, $N = 10$ и $n = 4$ находим $G_{кр} = 0,502$. Поскольку $G_{набл} = 0,3996 < G_{кр} = 0,502$, то можно считать точность автоматов практически одинаковой.

Проверка гипотезы о случайности и независимости результатов измерений в выборке. До проведения статистической обработки результатов измерения необходимо убедиться в том, что эти результаты являются стохастически (т.е. случайно) независимыми. Альтернативной гипотезой может быть предположение о наличии монотонного или циклического смещения (дрейфа) значения клика, вызванного неконтролируемым фактором. Подобный случай может иметь место при анализе размеров деталей, обрабатываемых на настроенном станке, когда вследствие изнашивания инструмента или нагрева станка центр группирования размеров постепенно смещается при неизменной стандартной погрешности. Наиболее мощным критерием проверки нулевой гипотезы будет критерий последовательных разностей τ .

Наблюдаемое значение критерия

$$\tau_{набл} = c^2 / s^2,$$

где

$$c^2 = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} (y_i - \bar{y})^2;$$

N – объем выборки; i – порядковый номер измерения отклика в выборке; s^2 – оценка дисперсии.

Критическое значение $\tau_{кр}$ определяется из таблицы критических значений критерия последовательных разностей в зависимости от принятой доверительной вероятности P и объема выборки N . Если $\tau_{набл} < \tau_{кр}$, то гипотеза о независимости и случайности измерений в выборке отвергается.

Сравнение двух дисперсий нормальных генеральных совокупностей. На практике задача сравнения дисперсий возникает, если требуется сравнить точность приборов, инструментов, самих методов измерений и т. д. Очевидно, что предпочтительнее использование тех приборов, инструментов или методов,

которые обеспечивают наименьшее рассеяние результатов измерений, т.е. наименьшую дисперсию.

Пусть генеральные совокупности X и Y распределены нормально. По независимым выборкам с объемами, соответственно равными n_1 и n_2 , извлеченными из этих совокупностей, найдены выборочные дисперсии s_x^2 и s_y^2 . Требуется по выборочным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные дисперсии рассматриваемых совокупностей равны между собой:

$$H_0: D(X) = D(Y).$$

Учитывая, что выборочные дисперсии являются несмещенными оценками генеральных дисперсий, т.е.

$$M[s_x^2] = D(X); M[s_y^2] = D(Y),$$

нулевую гипотезу можно записать так:

$$H_0: M[s_x^2] = M[s_y^2].$$

Таким образом, требуется проверить, равны ли математические ожидания выборочных дисперсий между собой. Такая задача ставится потому, что обычно выборочные дисперсии оказываются различными. Возникает вопрос: значимо (существенно) или незначимо различаются выборочные дисперсии?

Если окажется, что нулевая гипотеза справедлива, т.е. генеральные дисперсии одинаковы, то различие выборочных дисперсий незначимо и объясняется случайными причинами, в частности, случайным отбором объектов выборки. Например, если различие исправленных выборочных дисперсий результатов измерений, выполненных двумя приборами, оказалось незначимым, то приборы имеют одинаковую точность.

Если нулевая гипотеза отвергнута, т.е. генеральные дисперсии неодинаковы, то различие выборочных дисперсий значимо и не может быть объяснено случайными причинами, а является следствием того, что сами генеральные дисперсии различны. Например, если различие выборочных дисперсий результатов измерений, произведенных двумя приборами, оказалось значимым, то точность приборов различна.

В качестве критерия проверки нулевой гипотезы о равенстве генеральных дисперсий принимается отношение большей выборочной дисперсии к меньшей, т.е. случайная величина

$$F = s_{\sigma}^2 / s_M^2.$$

Величина F при условии справедливости нулевой гипотезы имеет распределение Фишера – Снедекора со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, где n_1 – объем выборки, по которой вычислена большая дисперсия, n_2 – объем выборки, по которой найдена меньшая дисперсия.

Критическая область строится в зависимости от вида конкурирующей гипотезы.

Первый случай

Нулевая гипотеза $H_0: D(X) = D(Y)$.

Конкурирующая гипотеза $H_1: D(X) > D(Y)$.

В этом случае строят одностороннюю, а именно – правостороннюю критическую область, исходя из требования, согласно которому вероятность попадания критерия F в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости:

$$P[F > F_{кр}(\alpha); k_1, k_2,] = \alpha.$$

Критическую точку $F_{кр}(\alpha; k_1, k_2)$ находят по таблице критических точек распределения Фишера – Снедекора, и тогда правосторонняя критическая область определяется неравенством $F > F_{кр}$, а область принятия нулевой гипотезы – неравенством $F < F_{кр}$.

Обозначим отношение большей выборочной дисперсии к меньшей, вычисленное по данным наблюдений, через $F_{набл}$ и сформулируем правила проверки нулевой гипотезы.

Для того чтобы при заданном уровне значимости проверить нулевую гипотезу $H_0: D(X) = D(Y)$ о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе $H_1: D(X) > D(Y)$, надо вычислить отношение большей дисперсии к меньшей, т.е.

$$F_{\text{набл}} = s_{\sigma}^2 / s_M^2,$$

и по таблице критических точек распределения Фишера – Снедекора, по заданному уровню значимости α и числам степеней свободы k_1 и k_2 (k_1 – число степеней свободы большей дисперсии) найти критическую точку $F_{\text{набл}}$ ($\alpha; k_1, k_2$).

Если $F_{\text{набл}} < F_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу. Если $F_{\text{набл}} > F_{\text{кр}}$, то нулевую гипотезу отвергают.

Пример. По двум независимым выборкам объемом $n_1 = 12$ и $n_2 = 15$, извлеченным из нормальных генеральных совокупностей X и Y , найдены исправленные выборочные дисперсии $s_x^2 = 11,41$ и $s_y^2 = 6,52$. При уровне значимости 0,05 проверить нулевую гипотезу $H_0: D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(X) > D(Y)$.

Решение. Найдем отношение большей дисперсии к меньшей:

$$F_{\text{набл}} = \frac{11,41}{6,52} = 1,75.$$

Конкурирующая гипотеза имеет вид $D(X) > D(Y)$, поэтому критическая область – правосторонняя. В таблице критических точек распределения Фишера по уровню значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = 12 - 1 = 11$ и $k_2 = 15 - 1 = 14$ находим критическую точку

$$F_{\text{кр}}(0,05; 11, 14) = 2,56.$$

Так как $F_{\text{набл}} < F_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные дисперсии различаются незначимо. Например, если бы рассматриваемые дисперсии характеризовали точность двух методов измерения, то следует признать равносильность сравниваемых методов.

Второй случай

Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу о равенстве генеральных дисперсий нормально распределенных совокупностей при конкурирующей гипотезе $H_1: D(X) \neq D(Y)$, надо вычислить отношение большей дисперсии к меньшей, т.е.

$$F_{\text{набл}} = s_{\sigma}^2 / s_M^2,$$

и в таблице критических точек распределения Фишера по уровню значимости $\alpha/2$ (вдвое меньшему заданного) и числам степеней свободы k_1 и k_2 (k_1 – число степеней свободы большей дисперсии) найти критическую точку $F_{\text{набл}}(\alpha/2; k_1, k_2)$. Если $F_{\text{набл}} < F_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу. Если $F_{\text{набл}} > F_{\text{кр}}$, то нулевую гипотезу отвергают.

Сравнение исправленной выборочной дисперсии с гипотетической генеральной дисперсией нормальной совокупности.

1. Дисперсия генеральной совокупности известна. Пусть генеральная совокупность распределена нормально, причем генеральная дисперсия хотя и неизвестна, но имеются основания предполагать, что она равна гипотетическому (предполагаемому) значению σ_0^2 . На практике σ_0^2 устанавливается на основании предшествующего опыта или теоретически.

Пусть из генеральной совокупности извлечена выборка объема n и по ней найдена выборочная дисперсия s^2 , имеющая $k = n - 1$ степеней свободы. Требуется по дисперсии при заданном уровне значимости проверить нулевую гипотезу, состоящую в том, что генеральная дисперсия рассматриваемой совокупности равна гипотетическому значению σ_0^2 . Другими словами, требуется установить, значимо или незначимо различаются выборочная и гипотетическая генеральная дисперсии.

Учитывая, что s^2 является несмещенной оценкой генеральной дисперсии, нулевую гипотезу можно записать так:

$$H_0: M(s^2) = \sigma_0^2.$$

На практике рассматриваемая гипотеза проверяется, если нужно проверить точность приборов, инструментов, станков, методов исследования и устойчивость технологических процессов. Например, если известна допустимая характеристика рассеяния контролируемого размера деталей, изготавливаемых станком-автоматом, равная σ_0^2 , а дисперсия, найденная по выборке, окажется значимо больше σ_0^2 , то станок требует подналадки.

В качестве критерия проверки нулевой гипотезы примем случайную величину $(n - 1) s^2 / \sigma_0^2$. Эта величина случайная, потому что в разных опытах s^2 принимает различные, наперед неизвестные значения. Можно показать, что она имеет распределение χ^2 с $k = n - 1$ степенями свободы:

$$\chi^2 = (n - 1) s^2 / \sigma_0^2.$$

Критическая область строится в зависимости от вида конкурирующей гипотезы.

1. Если нулевая гипотеза $H_0: \sigma^2 = \sigma_0^2$, а конкурирующая гипотеза $H_1: \sigma^2 > \sigma_0^2$, то в этом случае строят правостороннюю критическую область исходя из требования, согласно которому вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости:

$$P[\chi^2 > \chi_{кр}^2(\alpha; k)] = \alpha.$$

Критическую точку $\chi_{кр}^2(\alpha; k)$ находят по таблице критических точек распределения Хи-квадрат. Правосторонняя критическая область определяется неравенством $\chi^2 > \chi_{кр}^2$, а область принятия нулевой гипотезы – неравенством $\chi^2 < \chi_{кр}^2$.

2. Если нулевая гипотеза $H_0: \sigma^2 = \sigma_0^2$, а конкурирующая гипотеза $H_1: \sigma^2 \neq \sigma_0^2$, то строят двустороннюю критическую область. При этом исходят из требования, согласно которому вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости α .

Критические точки – левую и правую границы критической области – находят, требуя, чтобы вероятность попадания критерия в каждый из двух интервалов критической области была равна $\alpha / 2$:

$$P[\chi^2 < \chi_{лев.кр}^2(\alpha/2; k)] = \alpha/2;$$

$$P[\chi^2 > \chi_{прав.кр}^2(\alpha/2; k)] = \alpha/2.$$

В таблице критических точек распределения χ^2 указаны лишь «правые» критические точки. Возникает кажущееся затруднение в отыскании «левой»

критической точки, которое легко преодолеть, если принять во внимание, что события $\chi^2 < \chi_{\text{лев.кр}}^2$ и $\chi^2 > \chi_{\text{прав.кр}}^2$ противоположны и, следовательно, сумма их вероятностей равна единице:

$$P(\chi^2 < \chi_{\text{лев.кр}}^2) + P(\chi^2 > \chi_{\text{прав.кр}}^2) = 1$$

Отсюда

$$P(\chi^2 < \chi_{\text{лев.кр}}^2) = 1 - P(\chi^2 > \chi_{\text{прав.кр}}^2) = 1 - (\alpha/2)$$

Таким образом, левую критическую точку можно искать как правую (и значит, ее можно найти по таблице), исходя из требования, согласно которому вероятность попадания критерия в интервал, расположенный правее этой точки, должна быть равна $(1 - \alpha/2)$.

3. Наконец, если при той же нулевой гипотезе конкурирующая гипотеза $H_1: \sigma^2 < \sigma_0^2$, то находят критическую точку $\chi_{\text{кр}}^2(1 - \alpha; k)$, и если $\chi_{\text{набл}}^2 > \chi_{\text{кр}}^2(1 - \alpha; k)$, то нет оснований отвергнуть нулевую гипотезу, если же $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2(1 - \alpha; k)$, то нулевую гипотезу отвергают.

Пример. Из нормальной генеральной совокупности извлечена выборка объема $n = 13$, и по ней найдена исправленная выборочная дисперсия $s^2 = 14,6$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0: \sigma^2 = \sigma_0^2 = 12$, приняв в качестве конкурирующей гипотезы $H_1: \sigma^2 > 12$.

Решение. Определим наблюдаемое значение критерия:

$$\chi_{\text{набл}}^2 = (n - 1) s^2 / \sigma_0^2 = (13 - 1) \cdot 14,6 / 12 = 14,6.$$

По условию, конкурирующая гипотеза имеет вид $H_1: \sigma^2 > 12$, поэтому критическая область – правосторонняя.

По таблице критических точек распределения Хи-квадрат по уровню значимости 0,01 и числу степеней свободы $k = n - 1 = 13 - 1 = 12$ находим критическую точку $\chi_{\text{кр}}^2(0,01; 12) = 26,2$.

Так как $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2$, то нет оснований отвергнуть нулевую гипотезу. Другими словами, различие между исправленной дисперсией (14,6) и гипотетической генеральной дисперсией (12) незначимое.

Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых известны (независимые выборки). Пусть генеральные совокупности X и Y распределены нормально, причем их дисперсии известны (например, из предшествующего опыта или найдены теоретически). По независимым выборкам, объемы которых соответственно равны n и m , извлеченным из этих совокупностей, найдены выборочные средние \bar{X} и \bar{Y} .

Требуется по выборочным средним при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные средние (математические ожидания) рассматриваемых совокупностей равны между собой, т.е.

$$H_0: M(X) = M(Y).$$

Учитывая, что выборочные средние являются несмещенными оценками генеральных средних, т.е. $M(\bar{X}) = M(X)$ и $M(\bar{Y}) = M(Y)$, нулевую гипотезу можно записать так:

$$H_0: M(\bar{X}) = M(\bar{Y}).$$

Таким образом, требуется проверить, что математические ожидания выборочных средних равны между собой. Такая задача ставится потому, что, как правило, выборочные средние оказываются различными. Возникает вопрос: значимо или незначимо различаются выборочные средние?

Если окажется, что нулевая гипотеза справедлива, т.е. генеральные средние одинаковы, то различие выборочных средних незначимо и объясняется случайными причинами и, в частности, случайным отбором объектов выборки. Например, если физические величины A и B имеют одинаковые истинные размеры, а средние арифметические \bar{X} и \bar{Y} результатов измерений этих величин различны, то это различие незначимое.

Если нулевая гипотеза отвергнута, т.е. генеральные средние неодинаковы, то различие выборочных средних значимо и не может быть объяснено случайными причинами, а объясняется тем, что сами генеральные средние (математические ожидания) различны. Например, если среднее арифметическое \bar{X} результатов измерений физической величины A значимо отличается от

среднего арифметического \bar{y} результатов измерений физической величины В, то это означает, что истинные размеры (математические ожидания) этих величин различны.

В качестве критерия проверки нулевой гипотезы примем случайную величину

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{\sqrt{D(\bar{X})/n + D(\bar{Y})/m}},$$

которая является случайной, потому что в различных опытах \bar{X} и \bar{Y} принимают различные, наперед неизвестные значения.

Критерий Z – нормированная нормальная случайная величина. Действительно, величина Z распределена нормально, так как является линейной комбинацией нормально распределенных величин \bar{X} и \bar{Y} . В свою очередь, эти величины распределены нормально как выборочные средние, найденные по выборкам, извлеченным из нормальных генеральных совокупностей. Критерий Z является также нормированной величиной, потому что $M(Z) = 0$ при справедливости нулевой гипотезы $\sigma(Z) = 1$, поскольку выборки независимы.

Критическая область строится в зависимости от вида конкурирующей гипотезы.

1. Если нулевая гипотеза $H_0: M(\bar{X}) = M(\bar{Y})$, а конкурирующая гипотеза $H_1: M(\bar{X}) \neq M(\bar{Y})$, то строят двустороннюю критическую область. При этом исходят из требования, согласно которому вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости α . Наибольшая мощность критерия (вероятность попадания критерия в критическую область при справедливости конкурирующей гипотезы) достигается тогда, когда «левая» и «правая» критические точки выбраны так, что вероятность попадания критерия в каждый из двух интервалов критической области равна $\alpha/2$:

$$P(Z < z_{\text{лев.кр}}) = \alpha/2; \quad P(Z < z_{\text{прав.кр}}) = \alpha/2.$$

Поскольку Z – нормированная нормальная величина, а распределение такой величины симметрично относительно нуля, то критические точки

симметричны относительно нуля. Таким образом, если обозначить правую границу двусторонней критической области через $z_{кр}$, то левая граница будет равна $-z_{кр}$. Иными словами, достаточно найти правую границу, чтобы найти саму двустороннюю критическую область $Z < -z_{кр}$, $Z > z_{кр}$ и область принятия нулевой гипотезы $(-z_{кр}; z_{кр})$.

Правую границу двусторонней критической области находят, пользуясь функцией Лапласа $\Phi(z)$. Для этого необходимо найти значение аргумента функции Лапласа, которому соответствует значение функции, равное $(1 - 2\alpha)/2$. Тогда двусторонняя критическая область определяется неравенствами

$$Z < -z_{кр}, Z > z_{кр}$$

или равносильным неравенством $|Z| > z_{кр}$, а область принятия нулевой гипотезы – неравенством $-z_{кр} < Z < z_{кр}$ или равносильным неравенством $|Z| < z_{кр}$.

2. Если нулевая гипотеза $H_0: M(\bar{X}) = M(\bar{Y})$, а конкурирующая гипотеза $H_1: M(X) > M(Y)$, то строят правостороннюю критическую область. При этом исходят из требования, чтобы вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы была равна принятому уровню значимости:

$$P(Z > z_{кр}) = \alpha.$$

На практике такой случай имеет место, если профессиональные соображения позволяют предположить, что генеральная средняя одной совокупности больше генеральной средней другой. Например, если произведено усовершенствование конструкции режущего инструмента, то естественно допустить, что оно приведет к увеличению стойкости режущей пластины.

В рассматриваемом случае для того, чтобы найти границу правосторонней критической области ($z_{кр}$), достаточно определить значение аргумента функции Лапласа, которому соответствует значение функции, равное $(1 - 2\alpha)/2$. Тогда правосторонняя критическая область определяется неравенством $Z > z_{кр}$, область принятия нулевой гипотезы – неравенством Z

$< c$. Другими словами, если $Z_{\text{набл}} < z_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу. Если же $Z_{\text{набл}} > z_{\text{кр}}$, то нулевую гипотезу отвергают.

Пример. По двум независимым выборкам, объемы которых соответственно равны $n = 10$ и $m = 10$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние $\bar{x} = 14,3$ и $\bar{y} = 12,2$. Генеральные дисперсии известны: $D(X) = 22$, $D(Y) = 18$. При уровне значимости $0,05$ проверить нулевую гипотезу $H_0: M(\bar{X}) = M(\bar{Y})$ при конкурирующей гипотезе $H_1: M(X) > M(Y)$.

Решение. Определим наблюдаемое значение критерия:

$$Z_{\text{набл}} = \frac{14,3 - 12,2}{\sqrt{22/10 + 18/10}} = 1,05.$$

По условию конкурирующая гипотеза имеет вид $M(X) > M(Y)$, поэтому критическая область – правосторонняя.

По таблице функции Лапласа находим $z_{\text{кр}} = 1,64$.

Так как $Z_{\text{набл}} < z_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу. Другими словами, выборочные средние различаются незначимо.

3. Наконец, в том случае, если $H_0: M(X) > M(Y)$ при конкурирующей гипотезе $H_1: M(X) < M(Y)$, то используют тот факт, что критерий Z распределен симметрично относительно нуля. Поэтому искомая критическая точка $z_{\text{кр}}^{\cdot}$ симметрична такой точке $z_{\text{кр}} > 0$, для которой $P(Z > z_{\text{кр}}) = \alpha$, т.е. $z_{\text{кр}}^{\cdot} = -z_{\text{кр}}$. Таким образом, для того чтобы найти точку $z_{\text{кр}}^{\cdot}$, достаточно сначала найти «вспомогательную» точку $z_{\text{кр}}$ так, как это описано в предыдущем случае, а затем взять найденное значение со знаком минус. Тогда левосторонняя критическая область определяется неравенством $Z < -z_{\text{кр}}$, а область принятия нулевой гипотезы – неравенством $Z > -z_{\text{кр}}$. Поэтому, если $Z_{\text{набл}} > -z_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу, а если $Z_{\text{набл}} < -z_{\text{кр}}$, то нулевую гипотезу отвергают.

Сравнение двух средних произвольно распределенных генеральных совокупностей (большие независимые выборки). Метод сравнения двух средних, описанный выше, применим только в том случае, когда генеральные совокупности X и Y распределены нормально, а их дисперсии известны.

Однако если независимые выборки имеют большой объем (не менее 30 каждая), то указанное требование не является обязательным, поскольку при больших выборках выборочные средние распределены приближенно нормально, а выборочные дисперсии являются достаточно хорошими оценками генеральных дисперсий и в этом смысле их можно считать известными приближенно. В итоге критерий

$$Z' = \frac{\bar{X} - \bar{Y}}{\sqrt{D_B(X)/n + D_B(Y)/m}}$$

распределен приближенно нормально с параметрами $M(Z') = 0$ (при условии справедливости нулевой гипотезы) и $\sigma^2(Z') = 1$ (если выборки независимы). Поэтому можно сравнивать средние так, как описано в предыдущем разделе, только заменив точный критерий Z приближенным критерием Z' . В этом случае наблюдаемое значение приближенного критерия таково:

$$Z'_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{D_B(X)/n + D_B(Y)/m}}.$$

Однако, поскольку рассматриваемый критерий – приближенный, к выводам, полученным по этому критерию, следует относиться достаточно осторожно.

Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых неизвестны и одинаковы (малые независимые выборки). Предположим, что генеральные совокупности X и Y распределены нормально, причем их дисперсии неизвестны. Например, по выборкам малого объема нельзя получить хорошие оценки генеральных дисперсий. По этой причине метод сравнения средних, изложенный выше, применить нельзя.

Однако если дополнительно предположить, что эти неизвестные генеральные дисперсии равны между собой, то можно применить критерий

Стьюдента сравнения средних. Например, если сравниваются средние размеры двух партий деталей, изготовленных на одном и том же станке, то естественно допустить, что дисперсии контролируемых размеров одинаковы.

Если же нет оснований считать дисперсии одинаковыми, то, прежде чем сравнивать средние, следует, пользуясь описанным ранее критерием Фишера – Снедекора, предварительно проверить гипотезу о равенстве генеральных дисперсий.

Итак, предположим, что генеральные дисперсии оказались одинаковыми и можно проводить проверку нулевой гипотезы $H_0: M(X) = M(Y)$. Другими словами, требуется установить, значимо или незначимо различаются выборочные средние \bar{X} и \bar{Y} , найденные по независимым малым выборкам объемов n и m .

В качестве критерия проверки нулевой гипотезы примем случайную величину

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Известно, что величина T при справедливости нулевой гипотезы имеет t -распределение с $k = n + m - 2$ степенями свободы. Критическая область строится в зависимости от вида конкурирующей гипотезы.

1. Если нулевая гипотеза $H_0: M(X) = M(Y)$, а конкурирующая гипотеза $H_1: M(X) \neq M(Y)$, то строят двустороннюю критическую область. При этом исходят из требования, согласно которому вероятность попадания критерия T в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости α .

Наибольшая мощность критерия (вероятность попадания критерия в критическую область при справедливости конкурирующей гипотезы) достигается тогда, когда «левая» и «правая» критические точки выбраны так, что вероятность попадания критерия в каждый из двух интервалов двусторонней критической области равна $\alpha/2$:

$$P(T < t_{\text{лев.кр}}) = \alpha/2; P(T > t_{\text{прав.кр}}) = \alpha/2.$$

Поскольку величина T имеет распределение Стьюдента (а оно симметрично относительно нуля), то и критические точки симметричны относительно нуля. Таким образом, если обозначить правую границу двусторонней критической области через $t_{\text{двуст.кр}}(\alpha; \kappa)$, то левая граница равна $-t_{\text{двуст.кр}}(\alpha; \kappa)$. Другими словами, достаточно найти правую границу двусторонней критической области, чтобы найти саму двустороннюю критическую область: $T < -t_{\text{двуст.кр}}(\alpha; \kappa)$, $T > t_{\text{двуст.кр}}(\alpha; \kappa)$ и область принятия нулевой гипотезы: $[-t_{\text{двуст.кр}}(\alpha; \kappa), t_{\text{двуст.кр}}(\alpha; \kappa)]$.

Таким образом, чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0: M(X) = M(Y)$ о равенстве математических ожиданий двух нормальных совокупностей с неизвестными, но одинаковыми дисперсиями (в случае независимых малых выборок) при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$, надо вычислить наблюдаемое значение критерия

$$T_{\text{набл}} = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}$$

и в таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $k = n + m - 2$ найти критическую точку $t_{\text{двуст.кр}}(\alpha; \kappa)$. Если $|T_{\text{набл}}| < t_{\text{двуст.кр}}(\alpha; \kappa)$, то отвергнуть нулевую гипотезу нет оснований. Но если $|T_{\text{набл}}| > t_{\text{двуст.кр}}(\alpha; \kappa)$, то нулевую гипотезу отвергают.

Пример. По двум независимым малым выборкам, объемы которых соответственно равны $n = 5$ и $m = 6$, извлеченным из нормальных генеральных совокупностей X и Y , найдены выборочные средние $\bar{x} = 3,3$ и $\bar{y} = 2,48$, а также исправленные дисперсии $S_X^2 = 0,25$ и $S_Y^2 = 0,108$. При уровне значимости $0,05$ проверить нулевую гипотезу $H_0: M(X) = M(Y)$ при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$.

Решение. Так как выборочные дисперсии различны, проверим предварительно нулевую гипотезу о равенстве генеральных дисперсий,

пользуясь критерием Фишера – Снедекора. Для этого найдем отношение большей исправленной дисперсии к меньшей:

$$F_{\text{набл}} = 0,25 / 0,108 = 2,31.$$

Дисперсия S_X^2 значительно больше дисперсии S_Y^2 , поэтому в качестве конкурирующей примем гипотезу $H_1: D(X) > D(Y)$. В этом случае критическая область – правосторонняя. В таблице критических значений Фишера – Снедекора по уровню значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = 5 - 1 = 4$, $k_2 = 6 - 1 = 5$ находим критическую точку $F_{\text{кр}}(0,05; 4; 5) = 5,19$.

Так как $F_{\text{набл}} < F_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий. Поскольку предположение о равенстве генеральных дисперсий выполняется, сравним средние.

Вычислим наблюдаемое значение критерия Стьюдента:

$$T_{\text{набл}} = \frac{\bar{X} - \bar{Y}}{\sqrt{nS_X^2 + mS_Y^2}} \sqrt{\frac{nm(n + m - 2)}{n + m}}.$$

Подставив числовые значения величин, входящих в эту формулу, получим $T_{\text{набл}} = 3,27$. По условию конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область – двусторонняя. По уровню значимости 0,05 и числу степеней свободы $k_1 = 5 + 6 - 2 = 9$ находим по таблице критических точек распределения Стьюдента критическую точку $t_{\text{двуст.кр}}(0,05; 9) = 2,26$.

Так как $T_{\text{набл}} > t_{\text{двуст.кр}}$, то нулевую гипотезу о равенстве генеральных средние отвергаем. Другими словами, выборочные средние различаются значимо.

Сравнение выборочной средней с гипотетической генеральной средней нормальной совокупности. Рассмотрим два случая.

1. Дисперсия генеральной совокупности известна. Пусть генеральная совокупность X распределена нормально, причем генеральная средняя α хотя и неизвестна, но имеются основания предполагать, что она равна

гипотетическому (предполагаемому) значению α_0 . Например, если X – совокупность размеров X_i партии деталей, изготовляемых станком-автоматом, то можно предположить, что генеральная средняя α этих размеров равна наладочному размеру α_0 . Чтобы проверить это предположение, находят выборочную среднюю \bar{X} и устанавливают, значимо или незначимо различаются \bar{X} и α_0 . Если различие окажется незначимым, то станок обеспечивает в среднем наладочный размер; если различие значимое, то станок требует подналадки.

Предположим, что дисперсия генеральной совокупности известна, например, из предшествующего опыта, или найдена теоретически, или вычислена по выборке большого объема (по большой выборке можно получить достаточно хорошую оценку дисперсии).

Пусть из нормальной генеральной совокупности извлечена выборка объема n и по ней найдена выборочная средняя \bar{X} , причем генеральная дисперсия σ^2 известна. Требуется по выборочной средней при заданном уровне значимости проверить нулевую гипотезу $H_0: \alpha = \alpha_0$, т.е. о равенстве генеральной средней α гипотетическому значению α_0 .

Учитывая, что выборочная средняя является несмещенной оценкой генеральной средней, т.е. $M(\bar{X}) = \alpha$, нулевую гипотезу можно записать так: $M(\bar{X}) = \alpha_0$. Таким образом, требуется проверить, равно ли математическое ожидание выборочной средней гипотетической генеральной средней. Другими словами, надо установить, значимо или незначимо различаются выборочная и генеральная средние.

В качестве критерия проверки нулевой гипотезы примем случайную величину

$$U = (\bar{X} - \alpha_0) / \sigma(\bar{X}) = (\bar{X} - \alpha_0) \sqrt{n} / \sigma,$$

которая распределена нормально, причем при справедливости нулевой гипотезы $M(U) = 0$, $\sigma(U) = 1$.

Чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0: \alpha = \alpha_0$ о равенстве генеральной средней α нормальной совокупности с

известной дисперсией σ^2 гипотетическому значению α_0 при конкурирующей гипотезе $H_1: \alpha \neq \alpha_0$, надо вычислить наблюдаемое значение критерия:

$$U_{\text{набл}} = (\bar{X} - \alpha_0) \sqrt{n} / \sigma$$

и по таблице функции Лапласа найти критическую точку двусторонней критической области по равенству

$$\Phi(U_{\text{кр}}) = (1 - \alpha) / 2.$$

Если $|U_{\text{набл}}| < U_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу. Если же $|U_{\text{набл}}| > U_{\text{кр}}$, то нулевую гипотезу отвергают.

2. Дисперсия генеральной совокупности неизвестна, например, в случае малых выборок. Тогда в качестве критерия проверки нулевой гипотезы принимают случайную величину

$$T = (\bar{X} - \alpha_0) \sqrt{n} / s,$$

где s – «исправленное» среднее квадратическое отклонение.

Величина T имеет распределение Стьюдента с $k = n - 1$ степенями свободы.

Критическая область строится в зависимости от вида конкурирующей гипотезы. Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0: \alpha = \alpha_0$ о равенстве неизвестной генеральной средней α (нормальной совокупности с неизвестной дисперсией) гипотетическому значению α_0 при конкурирующей гипотезе $H_1: \alpha \neq \alpha_0$, надо вычислить наблюдаемое значение критерия:

$$T_{\text{набл}} = (\bar{X} - \alpha_0) \sqrt{n} / s$$

и в таблице критических точек распределения Стьюдента по заданному уровню значимости α , помещенному в верхней строке таблицы, и числу степеней свободы $k = n - 1$ найти критическую точку $t_{\text{двуст.кр}}(\alpha; k)$. Если $|T_{\text{набл}}| < t_{\text{двуст.кр}}$, то нет оснований отвергнуть нулевую гипотезу. Если же $|T_{\text{набл}}| > t_{\text{двуст.кр}}$, то нулевую гипотезу отвергают.

Пример. По выборке объема $n = 20$, извлеченной из нормальной генеральной совокупности, найдены выборочная средняя $\bar{x} = 16$ и «исправленное» среднее квадратическое отклонение $s = 4,5$. Требуется при уровне значимости $0,05$ проверить нулевую гипотезу $H_0: \alpha = \alpha_0 = 15$ при конкурирующей гипотезе $H_1: \alpha \neq 15$.

Решение.

Вычислим наблюдаемое значение критерия:

$$T_{\text{набл}} = (\bar{x} - \alpha_0) \sqrt{n}/s = (16 - 15) \cdot \sqrt{20}/4,5 = 0,99.$$

По условию, конкурирующая гипотеза имеет вид $\alpha \neq \alpha_0$ поэтому критически область – двусторонняя. В таблице критических точек распределения Стьюдента по уровню значимости $\alpha = 0,05$, помещенному в верхней строке таблицы, и по числу степеней свободы $K = 20 - 1 = 19$ находим критическую точку $t_{\text{двуст.кр}}(0,05; 19) = 2,09$.

Так как $|T_{\text{набл}}| < t_{\text{двуст.кр}}$, то нет оснований, чтобы отвергнуть нулевую гипотезу. Таким образом, выборочная средняя незначимо отличается от гипотетической генеральной средней.

Сравнение нескольких дисперсий нормальных генеральных совокупностей по выборкам различного объема. Критерий Бартлетта. Положим, что генеральные совокупности X_1, X_2, \dots, X_i распределены нормально. Из этих совокупностей извлечены независимые выборки различных объемов n_1, n_1, \dots, n_i (некоторые объемы могут быть одинаковыми; если все выборки имеют одинаковый объем, то предпочтительнее пользоваться критерием Кохрена, который описан в следующем подразделе). По выборкам найдены исправленные выборочные дисперсии $s_1^2, s_2^2, \dots, s_i^2$. Требуется по исправленным выборочным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные дисперсии рассматриваемых совокупностей равны между собой:

$$H_0: D(X_1) = D(X_2) = \dots = D(X_i).$$

Другими словами, требуется установить, значимо или незначимо различаются исправленные выборочные дисперсии. Гипотеза о равенстве нескольких дисперсий называется гипотезой об однородности дисперсий. Числом степеней свободы дисперсии s_i^2 называют число $k_i = n_i - 1$, т.е. число, на единицу меньшее объема выборки, по которой вычислена дисперсия,

Обозначим через \bar{s}^2 среднюю арифметическую исправленных дисперсий, взвешенную по числам степеней свободы:

$$s^2 = \sum_{i=1}^i k_i s_i^2 / k; \quad k = \sum_{i=1}^i k_i .$$

В качестве критерия проверки нулевой гипотезы об однородности дисперсий принимают критерий Бартлетта, т.е. случайную величину

$$B = V/C,$$

где

$$V = 2,303 \left[k \log \bar{s}^2 - \sum_{i=1}^i k_i \log s_i^2 \right]; \quad C = 1 + \frac{1}{3(i-1)} \left[\sum_{i=1}^i \frac{1}{k_i} - \frac{1}{k} \right].$$

Бартлетт установил, что случайная величина B при условии справедливости нулевой гипотезы распределена приблизительно как χ_2^2 с $i - 1$ степенями свободы, если все $k_i > 2$. Учитывая, что $k_i = n_i - 1$, заключаем, что $n_i - 1 > 2$, или $n_i > 3$, т.е. объем каждой из выборок должен быть не меньше 4.

Строят правостороннюю критическую область, исходя из требования, согласно которому вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости:

$$P[B > \chi_{кр}^2(\alpha; i - 1)] = \alpha .$$

Критическую точку $\chi_{кр}^2(\alpha; i - 1)$ находят по таблице критических точек распределения χ^2 по уровню значимости α и числу степеней свободы $k = i - 1$. Тогда правосторонняя критическая область определяется неравенством $B > \chi_{кр}^2$, а область принятия гипотезы – неравенством $B < \chi_{кр}^2$.

Чтобы при заданном уровне значимости α проверить нулевую гипотезу об однородности дисперсий нормальных совокупностей, надо вычислить наблюдаемое значение критерия Бартлетта $V_{\text{набл}} = V_{\text{набл}}/C$ и по таблице критических точек распределения χ^2 найти критическую точку $\chi_{\text{кр}}^2(\alpha; i - 1)$. Если $V_{\text{набл}} < \chi_{\text{кр}}^2$, то нет оснований отвергнуть нулевую гипотезу. Если же $V_{\text{набл}} > \chi_{\text{кр}}^2$, то нулевую гипотезу отвергают.

Необходимо заметить, что сначала лучше найти V и сравнить с $\chi_{\text{кр}}^2$, так как если окажется, что $V < \chi_{\text{кр}}^2$, то C вычислять уже не нужно (так как C всегда больше 1). Если же $V > \chi_{\text{кр}}^2$, то придется вычислить и C , а затем сравнить V с $\chi_{\text{кр}}^2$.

Сравнение нескольких дисперсий нормальных генеральных совокупностей по выборкам одинакового объема. Критерий Кохрена. Положим, что генеральные совокупности X_1, X_2, \dots, X_i распределены нормально. Из этих совокупностей извлечены независимые выборки одинакового объема, и по выборкам найдены исправленные выборочные дисперсии $s_1^2, s_2^2, \dots, s_i^2$, все с одинаковым числом степеней свободы $k = n - 1$. Требуется по исправленным выборочным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные дисперсии рассматриваемых совокупностей равны между собой:

$$H_0: D(X_1) = D(X_2) = \dots = D(X_i).$$

Другими словами, требуется установить, значимо или незначимо различаются исправленные выборочные дисперсии.

В рассматриваемом случае выборок одинакового объема можно по критерию Фишера – Снедекора сравнить наибольшую и наименьшую дисперсии. Если окажется, что различие между ними незначимо, то подавно незначимо и различие между остальными дисперсиями. Такой метод носит название критерия Кохрена. В качестве критерия проверки нулевой гипотезы принимается отношение максимальной исправленной дисперсии к сумме всех исправленных дисперсий:

$$G = s_{\text{max}}^2 / (s_1^2, s_2^2, \dots, s_i^2).$$

Распределение этой случайной величины зависит только от числа степеней свободы $k = n - 1$ и количества выборок i .

Строят правостороннюю критическую область исходя из требования, согласно которому вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости:

$$P[G > G_{кр}(\alpha; k, i)] = \alpha .$$

Критическую точку $G_{кр}(\alpha; k, i)$ находят по таблице критических точек распределения Кохрена, правосторонняя критическая область определяется неравенством $G > G_{кр}$, а область принятия нулевой гипотезы – неравенством $G < G_{кр}$.

Чтобы при заданном уровне значимости α проверить гипотезу об однородности дисперсий нормально распределенных совокупностей, надо вычислить наблюдаемое значение критерия $G_{набл}$ и по таблице найти критическую точку. Если $G_{набл} < G_{кр}$, то нет оснований отвергнуть нулевую гипотезу, если же $G_{набл} > G_{кр}$, то нулевую гипотезу отвергают.

Проверка гипотезы о нормальном распределении генеральной совокупности. Критерий согласия Пирсона. Если закон распределения генеральной совокупности неизвестен, но есть основания предположить, что он имеет определенный вид (назовем его A), то проверяют нулевую гипотезу: генеральная совокупность распределена по закону A .

Проверка гипотезы о предполагаемом законе неизвестного распределения производится так же, как и проверка гипотезы о параметрах распределения, т.е. при помощи специально подобранной случайной величины – критерия согласия.

Критерием согласия называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения. Наиболее часто применяют критерий согласия К. Пирсона (Хи-квадрат). Рассмотрим применение критерия Пирсона к проверке гипотезы о нормальном

распределении генеральной совокупности (критерий аналогично применяется и для других распределений, в этом состоит его достоинство). С этой целью сравнивают эмпирические (наблюдаемые) и теоретические (вычисленные в предположении нормального распределения) частоты.

Обычно эмпирические и теоретические частоты различаются. Возможно, что расхождение случайно (незначимо) и объясняется либо малым числом наблюдений, либо способом их группировки, либо другими причинами. Возможно, что расхождение частот неслучайно (значимо) и объясняется тем, что теоретические частоты вычислены исходя из неверной гипотезы о нормальном распределении генеральной совокупности.

Критерий Пирсона, как и любой критерий, не доказывает справедливость гипотезы, а лишь устанавливает на принятом уровне значимости ее согласие или несогласие с данными наблюдений.

Итак, пусть по выборке объема n построена гистограмма и получено эмпирическое распределение:

варианты (x_i) $x_1, x_2, \dots, x_s;$

эмпирические частоты (n_i) $n_1, n_2, \dots, n_s.$

Допустим, что в предположении нормального распределения генеральной совокупности вычислены теоретические частоты n_i^* . Для вычисления теоретических частот по массиву экспериментальных данных весь интервал наблюдаемых значений X (выборки объема n) делят на s частичных интервалов (x_i, x_{i+1}) одинаковой длины. Затем находят варианты, т.е. середины частичных интервалов $x_i^* = (x_i + x_{i+1})/2$. В качестве частоты n_i варианты x_i^* принимают число вариант, которые попали в i -й интервал. В результате получают последовательность равноотстоящих вариантов и соответствующих им частот:

$$x_1^*, x_2^*, \dots, x_s^*;$$

$$n_1, n_1, \dots, n_s.$$

Понятно, что сумма n_i будет равна объему выборки, т.е. n .

Затем вычисляют выборочную среднюю \bar{X}^* и выборочное среднее квадратическое отклонение σ^* . Производят нормирование случайной величины

X , т.е. переходят к величине $Z = (X - \bar{x}^*)/\sigma^*$, после чего вычисляют концы интервалов (z_i, z_{i+1}) :

$$z_i = (x_i - \bar{x}^*)/\sigma^*; \quad z_{i+1} = (x_{i+1} - \bar{x}^*)/\sigma^*.$$

При этом наименьшее значение Z , т.е. z_i , полагают равным $-\infty$, а наибольшее, т.е. z_{i+1} , полагают равным $+\infty$.

После этого вычисляют теоретические вероятности попадания X в интервалы (x_i, x_{i+1}) по выражению

$$p_i = \Phi(z_{i+1}) - \Phi(z_i),$$

где $\Phi(z)$ – функция Лапласа.

В заключение расчетов находят искомые теоретические частоты:

$$n_i^* = np_i.$$

Теперь при уровне значимости α требуется проверить нулевую гипотезу H_0 : генеральная совокупность распределена нормально. В качестве критерия проверки нулевой гипотезы принимают случайную величину

$$\chi^2 = \sum (n_i - n_i^*)^2 / n_i^*.$$

Эта величина случайная, так как в различных опытах она принимает различные, заранее не известные значения. Чем меньше различаются эмпирические и теоретические частоты, тем меньше величина критерия, и следовательно, он в известной степени характеризует близость эмпирического и теоретического распределений.

Число степеней свободы находят по равенству $k = s - 1 - r$, где s – число групп (частичных интервалов) выборки; r – число параметров предполагаемого распределения, которые оценены по данным выборки. В частности, если предполагаемое распределение – нормальное, то оценивают два параметра (математическое ожидание и среднее квадратическое отклонение), поэтому $r = 2$ и число степеней свободы $k = s - 1 - r = s - 1 - 2 = s - 3$.

Поскольку односторонний критерий более «жестко» отвергает нулевую гипотезу, чем двусторонний, построим правостороннюю критическую область. При этом исходим из требования, согласно которому вероятность попадания

критерия в эту область в предположении справедливости нулевой гипотезы должна быть равна принятому уровню значимости α :

$$P[\chi^2 > \chi_{кр}^2(\alpha; k)] = \alpha.$$

Таким образом, правосторонняя критическая область определяется неравенством $\chi^2 > \chi_{кр}^2(\alpha; k)$, а область принятия нулевой гипотезы – неравенством $\chi^2 < \chi_{кр}^2(\alpha; k)$.

Чтобы при заданном уровне значимости проверить нулевую гипотезу H_0 : генеральная совокупность распределена нормально, надо сначала вычислить теоретические частоты, а затем наблюдаемое значение критерия:

$$\chi_{набл}^2 = \sum (n_i - n_i')^2 / n_i$$

и в таблице критических точек распределения χ^2 по заданному уровню значимости α и числу степеней свободы $k = s - 3$ найти критическую точку $\chi_{кр}^2(\alpha; k)$. Если $\chi_{набл}^2 < \chi_{кр}^2$, то нет оснований отвергнуть нулевую гипотезу, если же $\chi_{набл}^2 > \chi_{кр}^2$, то нулевую гипотезу отвергают.

Применимость критерия Пирсона ограничивается минимальным объемом выборки – он не должен быть менее 50, чтобы обеспечить возможность разбить экспериментальный массив не менее чем на 5-8 вариантов. Малочисленные группы следует объединять в одну, суммируя частоты.

Пример. При уровне значимости 0,05 проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты:

Эмпирические частоты.....6 13 38 74 106 85 30 14

Теоретические частоты.....3 14 42 82 99 76 30 13

Решение. Вычисляем $\chi_{набл}^2 = 7,19$. Определяем число степеней свободы, учитывая, что число групп выборки (число различных вариантов) $s = 8$: $k = 8 - 3 = 5$.

В таблице критических точек распределения Хи-квадрат по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 5$ находим $\chi_{кр}^2(0,05; 5) = 11,1$. Так как $\chi_{набл}^2 < \chi_{кр}^2$, то нет оснований отвергнуть нулевую гипотезу. Другими словами, расхождение эмпирических и теоретических частот незначимое.

Следовательно, данные наблюдений согласуются с гипотезой о нормальном распределении генеральной совокупности.

Контрольные вопросы

1. Что называется критической областью принятия гипотезы?
2. Что называется ошибкой первого рода при проверке статистических гипотез?
3. Что называется ошибкой второго рода при проверке статистических гипотез?
4. Каким образом можно одновременно уменьшить ошибки первого и второго рода?
5. При помощи какого критерия проверяется гипотеза о равенстве дисперсий независимых выборок?
6. Какой критерий нужно использовать при сравнении дисперсий, если выборки имеют разный объем?
7. Какой критерий нужно использовать при сравнении дисперсий, если выборки имеют одинаковый объем?

6. ВЫБОР МОДЕЛИ МНОГОФАКТОРНОГО РЕГРЕССИОННОГО АНАЛИЗА

6.1. Предпланирование эксперимента

Выбор модели является одной из самых трудных задач планирования эксперимента. Выбрать модель – это значит выбрать вид функции отклика, т.е. записать ее уравнение:

$$y = f(x_1, x_2, \dots, x_k),$$

где $x_1 - x_k$ – независимые факторы.

Трудность выбора модели заключается в том, что этот выбор следует сделать до начала проведения экспериментов, а у исследователя обычно нет априорной (т.е. до опыта) достаточной информации, чтобы уверенно выбрать вид функции. В ряде случаев удается использовать опубликованные результаты подобных экспериментов или теоретические разработки, которые могут дать ориентировочное представление об ожидаемом виде функции отклика. Однако для более аргументированного выбора модели следует провести процедуру *интерпретации модели* (так иногда называют предпланирование эксперимента).

Чаще всего процедура планирования эксперимента проводится в два этапа. Первый этап иногда называют поисковым. На этом этапе выбирают очень узкие пределы варьирования факторов. Это позволяет приблизить поверхность отклика к плоской поверхности и тем самым адекватно отразить ее линейной моделью, благодаря чему при расчете коэффициентов модели достаточное число данных получается при малом числе экспериментов.

Для наглядности рассмотрим геометрический аналог функции отклика – поверхность отклика для эксперимента с двумя факторами. Изобразим графически возможные состояния «черного ящика» с двумя входами. Для этого достаточно располагать плоскостью с обычной декартовой системой координат. По одной оси координат будем откладывать в некотором масштабе значения (уровни) одного фактора, а по другой оси – второго. Тогда каждому состоянию «черного ящика» будет соответствовать точка на плоскости. У каждого фактора

есть минимальное и максимальное возможные значения, между ними расположена область определения, в пределах которой он может изменяться либо непрерывно, либо дискретно.

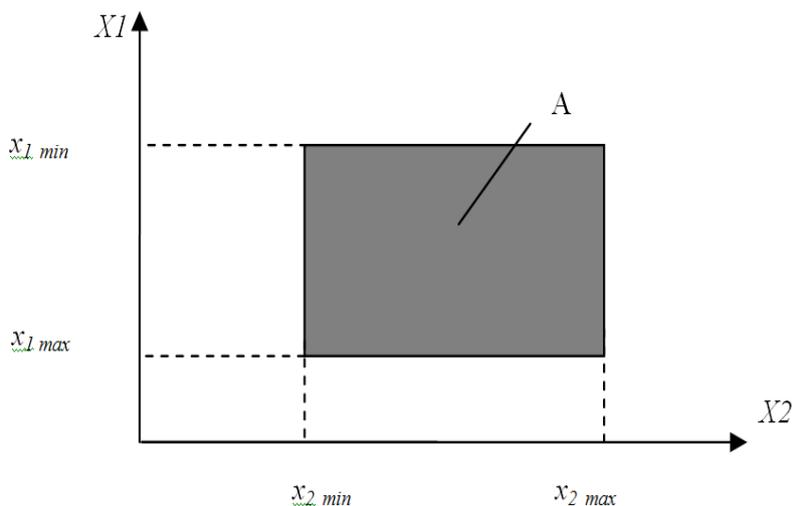


Рис. 6.1. Графическое представление возможных состояний «черного ящика»:

A – область определения факторов; B – поверхность отклика гипотетической модели

Если факторы совместимы, то границы образуют на плоскости некоторый прямоугольник, внутри которого лежат точки, соответствующие состояниям «черного ящика». Пунктирными линиями на рис. 6.1 обозначены границы областей определения каждого из факторов, а сплошными – границы их совместной области определения A .

Чтобы представить вид параметра оптимизации, требуется еще одна ось координат – обозначим ее U . В этом случае поверхность отклика B будет выглядеть так, как это показано на рис. 6.1. Пространство, в котором строится поверхность отклика, называют *факторным пространством*. Оно задается координатными осями, по которым откладываются значения факторов и параметра оптимизации. Размерность факторного пространства зависит от числа факторов. При многих факторах поверхность отклика уже нельзя изобразить наглядно и приходится ограничиваться только воображаемым гиперпространством и алгебраическими выражениями.

Для двух факторов наглядное представление можно получить путем сечения поверхности отклика плоскостями, параллельными плоскости X_1X_2 , и проецирования полученных в сечениях линий на эту плоскость. Получаются так называемые линии равных уровней, или *линии равного отклика* (рис. 6.2).

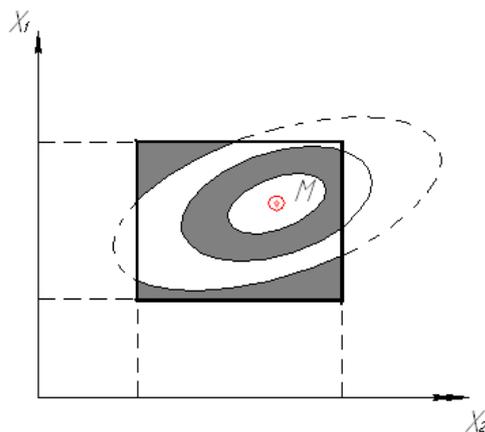


Рис. 6.2. Проекция на плоскость X_1X_2 сечений поверхности отклика:

М – точка оптимума

Каждая линия соответствует постоянному значению параметра оптимизации. Точка М на рисунке – это и есть оптимум, т.е. точка, в которой наблюдается максимальное значение параметра оптимизации. Существует соответствие между состоянием «черного ящика» и значением параметра оптимизации, но соответствие одностороннее, однонаправленное: каждому возможному состоянию «черного ящика» соответствует одно значение параметра оптимизации. Однако обратное утверждение неверно, поскольку одному возможному значению параметра оптимизации может соответствовать сколько угодно состояний.

В силу того, что невозможно измерить отклик в каждом состоянии, необходимо выбрать такую стратегию, которая приближала бы к оптимуму. Для этого используется так называемый *шаговый принцип предпланирования эксперимента*. Предположим, что оптимум существует и что поверхность отклика непрерывная, гладкая и имеет единственный оптимум. В большинстве случаев в области резания, металлорежущих станков и инструментов, и технологии машиностроения такое предположение, как показывает практика, справедливо. А если это так, то можно представить изучаемую функцию в виде

аналитической, в частности – в виде степенного ряда в окрестности любой возможной точки факторного пространства.

На рис. 6.3 показаны два варианта поиска оптимума. Точками на рисунке обозначены условия опытов. В варианте на рис. 6.3, *а* использован подход, который называют методом Гаусса – Зейделя.

Он состоит в том, что сначала последовательно изменяют значения одного фактора (эксперимент 1). Находят и фиксируют наилучшее значение этого фактора. Затем, не меняя наилучшее значение первого фактора, последовательно изменяют значения второго фактора (эксперимент 2). В случае, показанном на рис. 6.3, *б*, представлена шаговая процедура. Сначала изучается некоторая локальная область (1), в которой определяют направление максимального увеличения отклика при изменении параметра 2, а затем в этом направлении (1–2) ставят последующие опыты. Это схематичный пример, но он показывает определенные преимущества шаговой процедуры предпланирования.

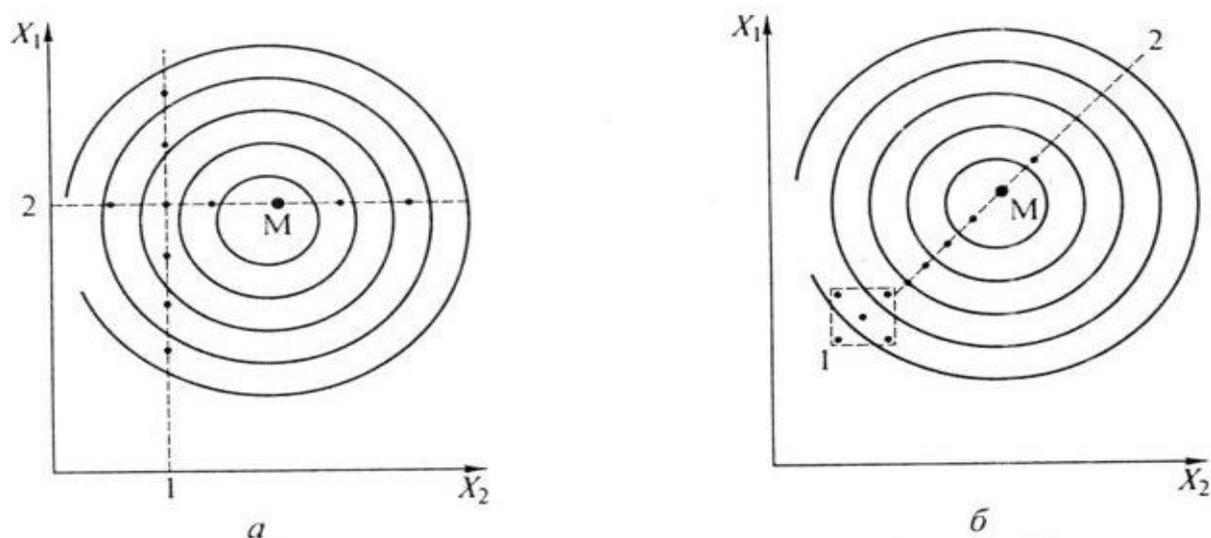


Рис. 6.3. Два варианта поиска оптимума (точки М):

а – последовательное изменение фактора X_1 (по направлению 1) и фактора X_2 (по направлению 2); *б* – движение по нормали к поверхности отклика (по направлению 1-2)

6.2. Шаговая процедура предпланирования

Реализацию шагового метода выполняют путем проведения серии однофакторных экспериментов по следующей методике. Например, необходимо разработать математическую модель для исследования амплитуды колебаний при резании:

$$A = f(s_0, t, v, \alpha),$$

где s_0 – подача (0,05...0,8 мм/об); t – глубина резания (1 ...4 мм); v – скорость резания (60... 180 м/мин), α – задний угол (3... 12°). Ясно, что это будет четырехфакторный эксперимент, но какого вида функция будет описывать параметр оптимизации, т.е. амплитуду колебаний, предположить очень сложно. Поэтому с целью получения предварительного представления о поверхности уровня проводят несколько однофакторных экспериментов [5, 12]. Например, вначале реализуют однофакторный эксперимент $A = f(s_0)$ при средних значениях остальных факторов ($t = 2,5$ мм, $v = 120$ м/мин, $\alpha = 7,5^\circ$), затем – однофакторный эксперимент вида $A = f(t)$ при условиях резания ($s_0 = 0,425$ мм, $v = 120$ м/мин, $\alpha = 7,5^\circ$). Третий и четвертый этапы проводят аналогично, т.е. их целью являются однофакторные эксперименты $A = f(v)$ и $A = f(\alpha)$ при установленных средних значениях остальных факторов. Затем строят однофакторные зависимости, по которым можно ориентировочно выбрать вид функции (на рис. 6.4 оказаны две из них). Приведенные зависимости явно нелинейным образом зависят от факторов. О нелинейности свидетельствует и характер сечений поверхности отклика плоскостями, параллельными координатной плоскости (s_0, α).

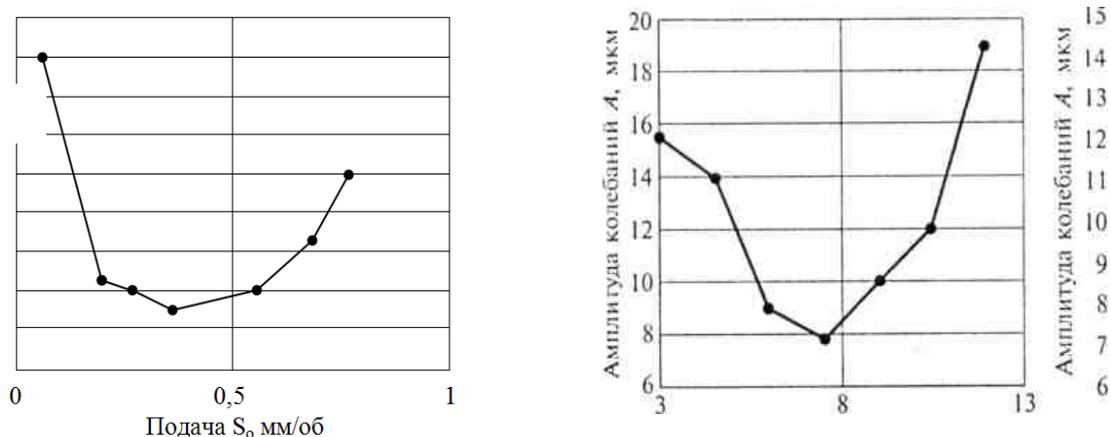


Рис. 6.4. Однофакторные зависимости амплитуды колебаний при резании от заднего угла $A=f(\alpha)$ (f) и от подачи $A=f(s_0)$ (б)

Если выбрать для первого шага предпланирования эксперимента область определения факторов $s_0 = 0,05 \dots 0,10$ мм/об и $\alpha = 3,0 \dots 4,5^\circ$ и провести эксперимент 2^2 , т.е. двухфакторный эксперимент на двух уровнях, то можно получить линейную модель, приближенно описывающую эту область поверхности.

Из модели следует, что при изменении подачи происходит более быстрое продвижение к оптимуму – минимальному значению амплитуды колебаний. Поэтому для второго шага выбираем следующую область определения факторов, смещая исходную область определения преимущественно в сторону увеличения фактора s_0 . В этой новой области определения факторов на основании новых экспериментов будет получена другая модель, в которой большее значение коэффициента будет уже при факторе α . Поэтому для третьего шага следует выбрать направление, в большей степени ориентированное на изменение заднего угла. Описанное пошаговое изменение направления выбора областей определения факторов показано на рис. 6.5 стрелками. Обобщая данный пример, можно заметить, что из полученных расчетом коэффициентов выбирают наибольший – именно он будет определять следующий шаг экспериментов, т.е. следующей области определения факторов. Такая последовательность экспериментов называется *движением по градиенту*.

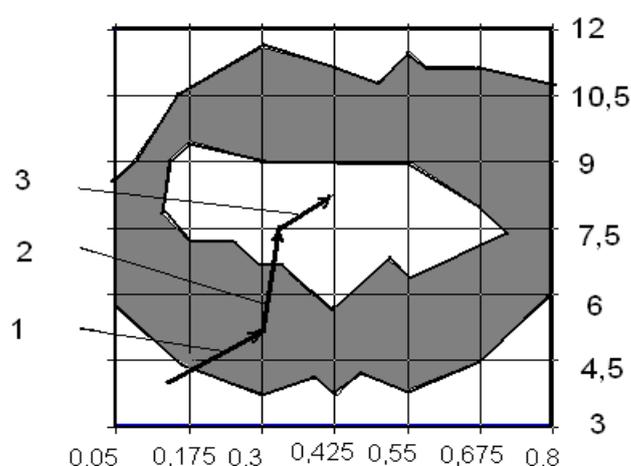


Рис. 6.5. Контурные области сечений поверхности отклика $A = f(\alpha, s_0)$ плоскостями, параллельными координатной плоскости (s_0, α) : 1, 2, 3 – последовательное движение по градиенту

Эти действия постепенно приближают исследователя к области оптимума, когда коэффициенты при факторах, постепенно уменьшаясь по абсолютной величине, начинают менять свой знак. Это означает, что предпланирование завершено и необходимо переходить к интерполяционной модели.

6.3. Интерполяционная модель

На этом, завершающем, этапе планирования в области оптимума, как правило, необходимо последовательно повышать степень полинома и рассчитывать коэффициенты модели до тех пор, пока не будет достигнута ее адекватность. Поэтому в планировании эксперимента общепринята такая процедура: сначала выбирают простейшую модель (обычно линейную, т.е. модель первого порядка) и находят для нее коэффициенты. Подставляют в эту модель значения параметров и рассчитывают отклик y . Затем по специальным формулам статистики проводят оценку адекватности модели (формулы рассматриваются далее). Если адекватность доказана, то эксперимент считают завершенным. Если же адекватности нет, берут следующий по порядку полином и снова после расчетов оценивают адекватность. Увеличение порядка полинома проводят до тех пор, пока не будет достигнута адекватность модели или, если число членов в усложненной полиномиальной модели превышает число экспериментов, проводят дополнительные эксперименты. В некоторых случаях, если проведение самих экспериментов не связано с большими расходами, а большую часть расходов составляет подготовка условий для проведения экспериментов, поступают в обратном порядке. Проводят избыточное число экспериментов, берут за основу модель высокого порядка и выполняют расчеты коэффициентов модели и оценку адекватности, поэтапно упрощая модель до тех пор, пока не будет достигнут предел точности получения экспериментальных данных.

Для описанной процедуры рекомендуется выбирать модель в виде степенных функций следующего вида (такого типа функции часто применяют при исследовании процессов резания):

$$y = C_0 s_0^{C_1} t^{C_2} v^{C_3} \alpha^{C_4}.$$

Приведенную модель иногда записывают в логарифмическом виде:

$$\ln y = \ln C_0 + C_1 \ln t + C_3 \ln v + C_4 \ln \alpha,$$

а после вычислений производят потенцирование. Это позволяет при расчете коэффициентов модели использовать приемы расчета, отработанные для полиномов. С помощью специальных формул, о которых будет идти речь далее, по результатам экспериментов рассчитывают постоянные коэффициенты C_0 , C_1 , C_2 , C_3 и C_4 .

Очень удобным видом модели являются *полиномиальные модели*, которые позволяют выявить влияние на параметр оптимизации не только отдельных факторов, но и их взаимодействие. Полиномиальные модели именуется по наибольшей степени фактора в полиноме:

– модель нулевой степени $y = b_0$;

– модель первой степени $y = b_0 + b_1 x_1 + b_2 x_2$;

– модель второй степени $y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2$;

– модель третьей степени $y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{112} x_1^2 x_2 + b_{112} x_1 x_2^2 + b_{111} x_1^3 + b_{222} x_2^3$.

Чем сложнее функция отклика, тем более высокий порядок полинома приходится использовать для достижения требуемой точности математического описания исследуемого процесса. Как правило, функция отклика нам неизвестна, но мы предполагаем (на основе теории, литературных данных, но лучше всего – на основе проведенных однофакторных экспериментов), что ее *приближенно* можно описать одной из представленных выше формул. Операция замены неизвестной функции по некоторой совокупности экспериментальных откликов другой функцией, условно говоря, эквивалентной, называется *аппроксимацией*, т.е. мы аппроксимируем неизвестную функцию полиномом. На рис. 6.6 представлен пример уменьшения погрешности аппроксимации экспериментальных данных по мере увеличения порядка полиномиальной модели. Как видим, чем больше членов в полиномиальной модели, тем более точно она отражает экспериментальные

результаты. Однако, как будет показано в дальнейшем изложении, с увеличением числа членов необходимо провести большее число экспериментов, чтобы найти значения коэффициентов.

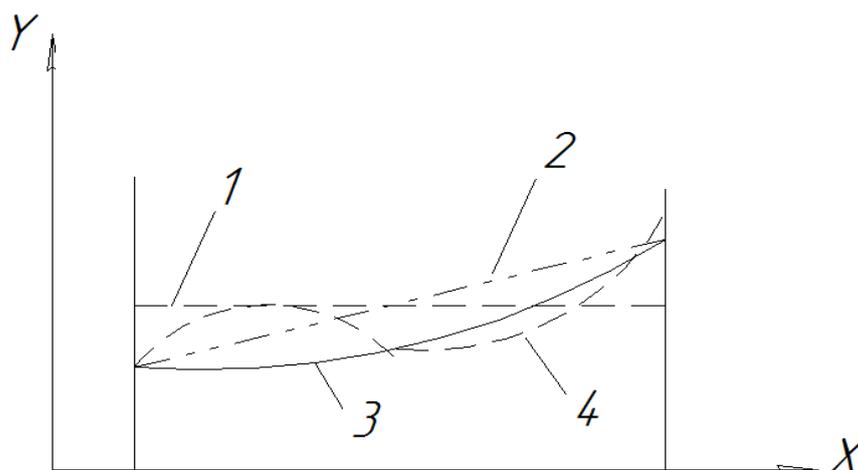


Рис. 6.6. Примеры аппроксимации экспериментальных точек (.) полиномами нулевого (1), первого (2), второго (3) и третьего (4) порядков

Необходимо отметить еще одно обстоятельство. Как ни странно это звучит, существует понятие «чрезмерной (или неоправданной) точности». Дело заключается в том, что при любом экспериментальном исследовании есть некоторая ошибка эксперимента, причины которой разнообразны: нестабильность условий (не поддающиеся регулированию колебания напряжения в сети, температуры в помещении и зависимости от времени суток и т.п.), разброс характеристик параметров факторов (ошибки установки глубины резания по лимбу, неточности заточки углов режущего клина и т.п.) и другие отклонения. Поэтому не имеет смысла усложнять модель сверх той, которая обеспечит точность аппроксимации выше точности эксперимента.

Точность аппроксимации оценивается по специальной методике, путем выдвижения и статистической оценки гипотезы об *адекватности* модели. Адекватность модели — это способность модели предсказывать результаты поведения объекта исследования в определенной области изменения факторов с заданной точностью.

Контрольные вопросы

1. В чем заключается предпланирование эксперимента?
2. Какие варианты поиска оптимума функции отклика существуют?
3. В чем заключается шаговая процедура предпланирования?
4. Что такое интерполяционная модель?
5. Дайте понятие адекватности модели.

7. ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ. ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ

Как при нахождении математической модели технологического процесса в заданной части факторного пространства, так и при нахождении оптимальных условий протекания процесса на первом этапе исследований требуется построение линейных моделей (см. рис. 7.1), а уже затем принимается решение о действиях на следующих этапах.

Для построения линейных моделей процессов широко применяются математические планы, соответствующие полному факторному эксперименту (ПФЭ) или дробному факторному эксперименту (ДФЭ).

Полным факторным экспериментом называется такой эксперимент, при реализации которого определяется значение параметра оптимизации при всех возможных сочетаниях уровней варьирования факторов. Если мы имеем дело с k -факторами, каждый из которых может устанавливаться на q -уровнях, то для того чтобы осуществить ПФЭ, необходимо поставить $n = q^k$ опытов.

Планирование, проведение и обработка результатов ПФЭ состоят из следующих обязательных этапов: кодирование факторов и определение интервалов их варьирования; составление плана-матрицы эксперимента; проверка воспроизводимости опытов; оценка значимости коэффициентов регрессии; проверка адекватности линейной модели.

Наибольшее распространение получили эксперименты, в которых факторы варьируются на двух уровнях (планы 2^k). Реже встречаются 3^k , так как с ростом числа уровней факторов возрастает количество опытов.

Кодирование факторов. Поскольку факторы изучаемого процесса неоднородны и имеют различные единицы измерения, их следует привести к единой системе исчисления путем перехода от действительных значений факторов к кодированным. Связь между кодовым и натуральным значениями фактора задается формулой

$$x_i = \frac{\tilde{x}_i - \tilde{x}_{i0}}{\delta_{\tilde{x}_i}};$$

где \tilde{x}_i – натуральное значение фактора;

\tilde{x}_{i0} – натуральное значение фактора на нулевом (основном) уровне,

$$\tilde{x}_{i0} = \frac{\tilde{x}_{i\max} + \tilde{x}_{i\min}}{2}; \quad \delta_{\tilde{x}_i} = \frac{\tilde{x}_{i\max} - \tilde{x}_{i\min}}{2};$$

x_i – кодированное значение фактора.

Вопрос о выборе интервала варьирования требует тщательного рассмотрения. В некоторых случаях границы области исследования факторного пространства могут совпадать с границами интервала варьирования. Однако, особенно при оптимизации процесса, вначале целесообразно описать его линейным уравнением, и поэтому интервал варьирования должен быть достаточно мал для получения линейного уравнения, но, вместе с тем, достаточно велик, чтобы не получить ошибочного вывода о незначимости какого-либо фактора. В табл. 7.1 приведены результаты кодирования факторов диффузионной сварки сплава ВЖЛ12У со сталью ЭИ961.

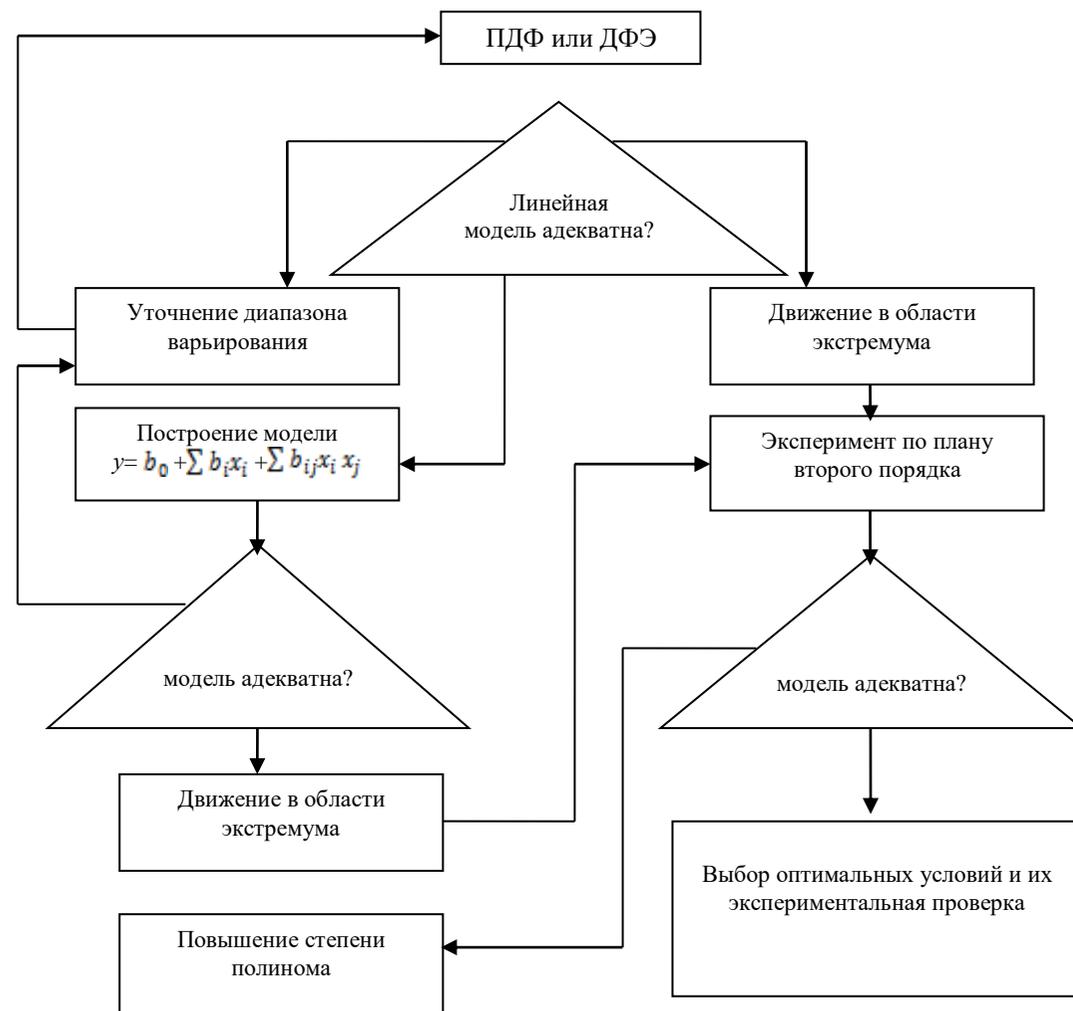
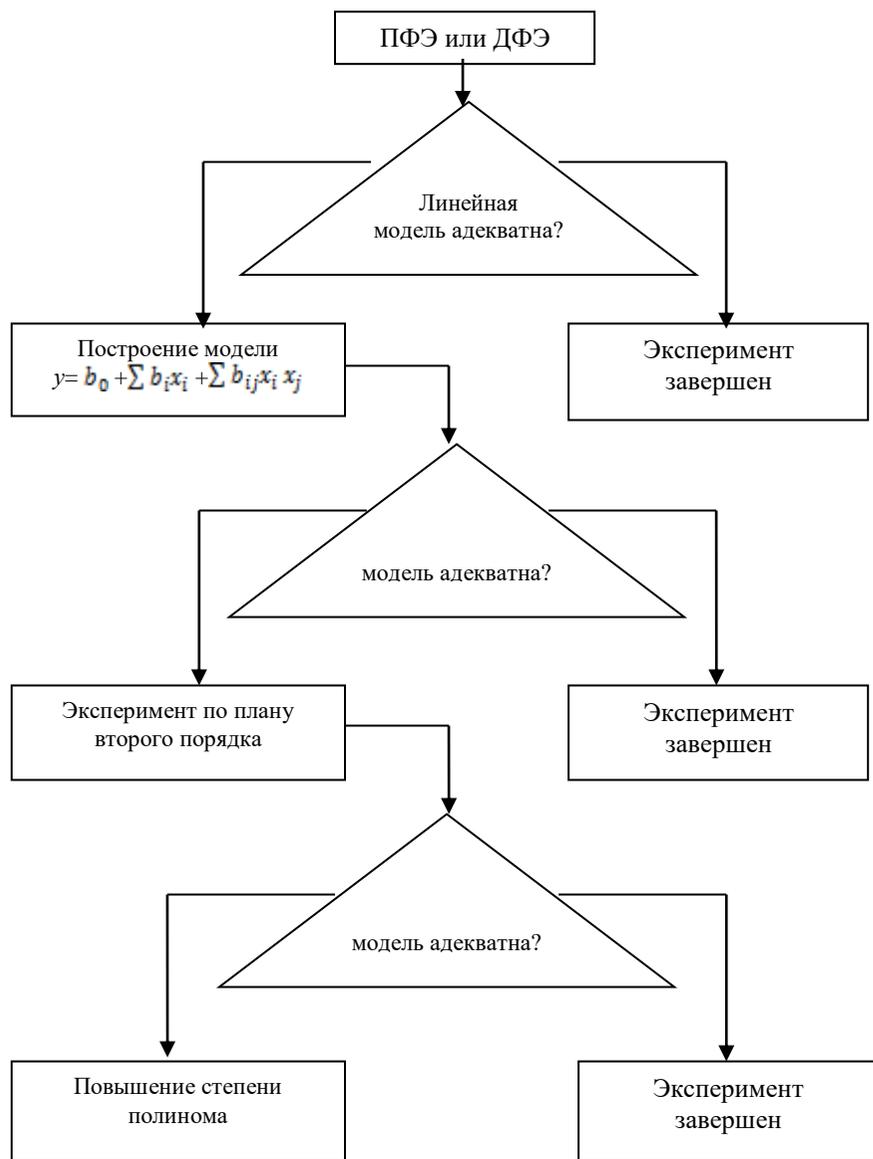


Рис. 7.1. Алгоритмы планирования полного (ПФЭ) или дробного (ДФЭ) факторных экспериментов

**Кодированные и натуральные значения факторов
процесса диффузионной сварки сплава ВЖЛ12У со сталью ЭИ961**

Параметры режима	Код	Нулевой уровень, 0	Интервал варьирования, δ_{x_i}	Нижний уровень, -1	Верхний уровень, +1
$T_{св}$, К	x_1	1348	55	1293	1403
$P_{св}$, МПа	x_2	12,5	4,5	8,0	17,0
$\tau_{св}$, мин	x_3	12,5	4,5	8,0	17,0

Составление план-матрицы эксперимента. Составление план-матрицы полного факторного эксперимента происходит следующим образом: для x_1 уровни чередуются в каждом опыте; для x_2 – через 2 опыта и т.д. В результате план-матрица содержит все возможные сочетания факторов исследуемого процесса. План-матрица для ПФЭ типа 2^3 приведена в табл. 7.2 и на рис. 7.2.

Планирование согласно ПФЭ называется ортогональным планированием, при котором определение коэффициентов уравнения регрессии осуществляется независимо друг от друга.

Таблица 7.2

Матрица планирования для ПФЭ 2^3

№ опыта	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	-1	-1	-1	+1	+1	+1	- +1
2	+1	-1	-1	-1	-1	+1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1

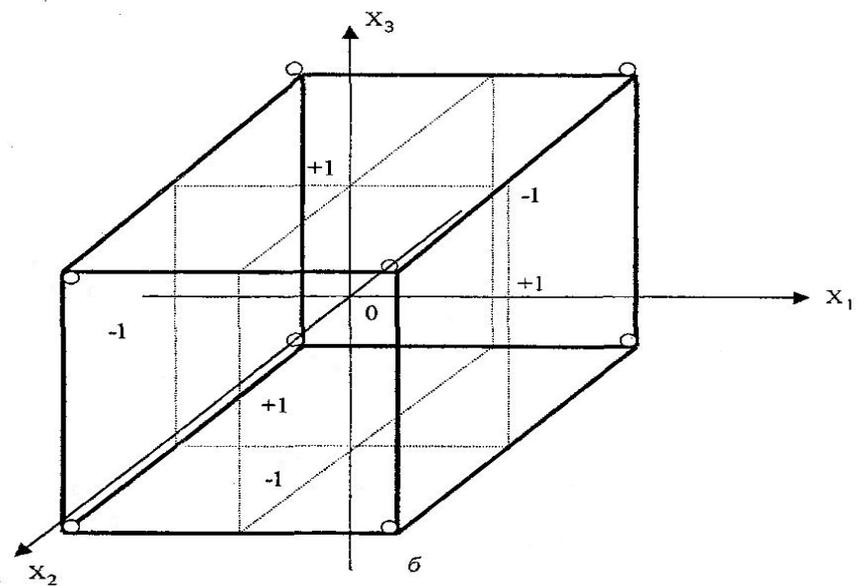
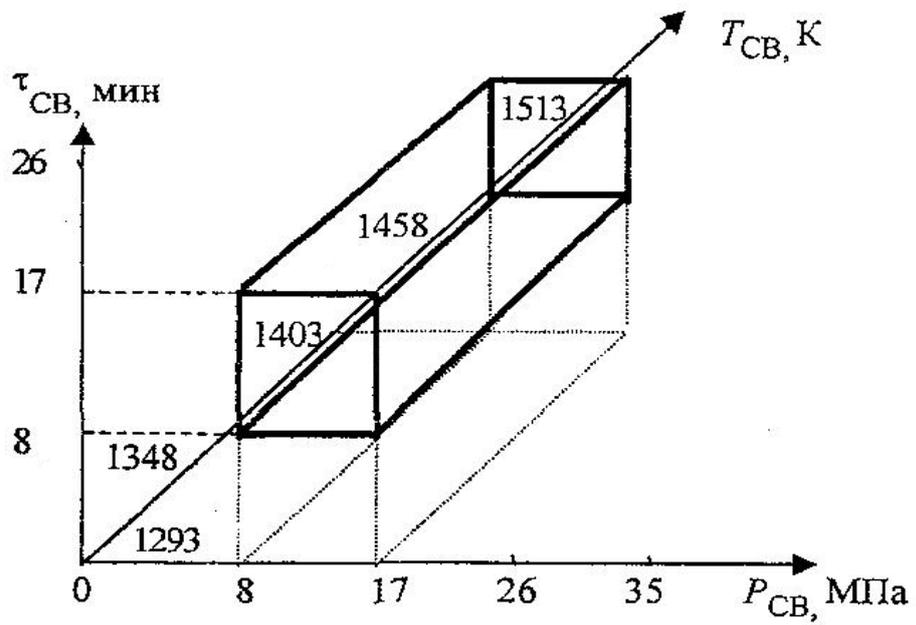


Рис. 7.2. Геометрическая интерпретация для плана ПФЭ 2^3 в натуральных (а) и кодовых (б) единицах измерений

Для составления плана эксперимента, который характеризуется четырьмя факторами, т.е. ПФЭ типа 2^4 , матрицу, приведенную в табл. 7.2, необходимо повторить дважды: при x_4 на нижнем и верхнем уровнях.

Планы факторного эксперимента могут быть записаны в компактной форме: вместо каждой строки в матрице выписать только те факторы, которые находятся на верхнем уровне, а строчку, в которой все факторы находятся на нижнем уровне, обозначить через 1 (табл. 7.3).

Рандомизация опытов. Рандомизацией называется процедура установления случайного порядка проведения опытов по времени, так как кроме варьируемых факторов при исследовании технологического процесса имеется целый ряд дополнительных факторов, оказывающих влияние на функцию отклика. Чтобы внести элемент случайности влияния дополнительных факторов на результат в целях обоснованного применения аппарата математической статистики, осуществляется рандомизация опытов. Для ее осуществления пользуются таблицами случайных чисел, извлечением номеров опытов из урны и т.п.

Реализация плана эксперимента. В результате реализации плана эксперимента определяются значения откликов или параметров оптимизации в каждом опыте и фиксируются в таблице результатов. В рассматриваемом примере в качестве параметров оптимизации были выбраны прочность сварного соединения на растяжение Y и остаточная макропластическая деформация Z . Для оценки воспроизводимости опытных данных план эксперимента был реализован дважды (табл. 7.4).

$$G_Y = \frac{S_{u_{max}}^2}{\sum_{u=1}^n S_u^2} = \frac{512}{1738} \approx 0,3 < G_{(0,05;8;1)} = 0,6798$$

$$G_Z = \frac{0,72}{2,465} \approx 0,3 < G_{(0,05;8;1)} = 0,6798;$$

$$S_Y^2 = \frac{\sum_{u=1}^n S_u^2}{n} = \frac{1738}{8} = 217,25;$$

$$S_Z^2 = \frac{2,465}{8} = 0,308.$$

Планируя эксперимент, мы стремимся получить линейную модель. Однако у нас нет уверенности в том, что в выбранных интервалах варьирования процесс описывается линейной моделью. Один из часто встречающихся видов нелинейности связан с тем, что влияние одного фактора на отклик зависит от

уровня, на котором находится другой фактор. Это означает, что имеется эффект взаимодействия факторов. Полный факторный эксперимент позволяет количественно оценивать эффект взаимодействия. Для этого необходимо воспользоваться расширенной план-матрицей, в которой учитывается произведение факторов.

Таблица 7.3

Матрицы для ПФЭ типов 2^2 , 2^3 , 2^4 , 2^5

Опыт	Тип эксперимента			
	2^2	2^3	2^4	2^5
1	I	I	I	I
2	x_1	x_1	x_1	x_1
3	x_2	x_2	x_2	x_2
4	x_1, x_2	x_1, x_2	x_1, x_2	x_1, x_2
5		x_3	x_3	x_3
6		x_1, x_3	x_1, x_3	x_1, x_3
7		x_2, x_3	x_2, x_3	x_2, x_3
8		x_1, x_2, x_3	x_1, x_2, x_3	x_1, x_2, x_3
9			x_4	x_4
10			x_1, x_4	x_1, x_4
11			x_2, x_4	x_2, x_4
12			x_1, x_2, x_4	x_1, x_2, x_4
13			x_3, x_4	x_3, x_4
14			x_1, x_3, x_4	x_1, x_3, x_4
15			x_2, x_3, x_4	x_2, x_3, x_4
16			x_1, x_2, x_3, x_4	x_1, x_2, x_3, x_4
17				x_5
18				x_1, x_5
19				x_2, x_5
20				x_1, x_2, x_5
21				x_3, x_5
22				x_1, x_3, x_5
23				x_2, x_3, x_5
24				x_1, x_2, x_3, x_5
25				x_4, x_5
26				x_1, x_4, x_5
27				x_2, x_4, x_5
28				x_1, x_3, x_4, x_5
29				x_3, x_4, x_5
30				x_1, x_3, x_4, x_5
31				x_2, x_3, x_4, x_5
32				x_1, x_2, x_3, x_4, x_5

Таблица 7.4

Результаты экспериментов и оценка их воспроизводимости

Опыт	x_1	x_2	x_3	x_1*x_2	x_1*x_3	$x_1*x_2*x_3$	x_2*x_3	Y_{u1} , МПа	Y_{u2} , МПа	\bar{Y}_u , МПа	$S^2_{Y_u}$	Z_{u1} , %	Z_{u2} , %	\bar{Z}_u , %	$S^2_{Z_u}$
1	-1	-1	-1	+1	-1	-1	+1	690	720	705	450	0,65	0,95	0,8	0,045
2	+1	-1	-1	-1	-1	+1	+1	1015	1025	1020	50	2,9	3,7	3,3	0,32
3	-1	+1	-1	-1	+1	+1	-1	895	915	905	200	1,3	2,5	1,9	0,72
4	+1	+1	-1	+1	-1	-1	-1	998	1022	1010	288	6,8	7,4	7,1	0,18
5	-1	-1	+1	+1	-1	+1	-1	849	863	856	78	0,7	1,3	1,0	0,18
6	+1	-1	+1	-1	+1	-1	-1	1029	1061	1045	512	4,4	4,6	4,5	0,02
7	-1	+1	+1	-1	-1	-1	+1	929	937	933	32	2,4	3,4	2,9	0,50
8	+1	+1	+1	+1	+1	+1	+1	1002	1018	1010	128	8,6	9,6	9,1	0,50

При этом математическая модель процесса ДСВ сплава ВЖЛ12У со сталью ЭИ961 будет иметь вид системы двух уравнений регрессии:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3;$$

$$Z = b'_0 + b'_1x_1 + b'_2x_2 + b'_3x_3 + b'_{12}x_1x_2 + b'_{13}x_1x_3 + b'_{23}x_2x_3 + b'_{123}x_1x_2x_3.$$

Проверка воспроизводимости опытов. При одинаковом числе параллельных опытов на каждом сочетании уровней факторов воспроизводимость процесса проверяется по критерию Кохрена:

$$G = \frac{S^2_{u_{\max}}}{\sum_{u=1}^n S^2_u} \leq G(0,05; f_u; f_n), \quad (7.1)$$

где $S^2_{u_{\max}}$ – наибольшая дисперсия в строчках плана;

$$S^2_u = \frac{\sum_{p=1}^m (Y_{pu} - \bar{Y}_u)^2}{m-1} - \text{дисперсия, характеризующая рассеяние результатов}$$

опытов на u -том сочетании уровней факторов;

$\delta = 1, 2, \dots, m$ – число параллельных опытов;

$G(0,05; f_u; f_n)$ – табличное значение критерия Кохрена при 5%-ном уровне значимости;

$f_u = m - 1$ – число степеней свободы каждой оценки;

$f_n = n$ – число независимых оценок дисперсии;

Y_{pu} – результат отдельного опыта;

\bar{Y}_u – среднее значение опытов на u -том сочетании уровней факторов.

Число степеней свободы f – понятие, учитывающее в статистических ситуациях связи, ограничивающие свободу изменения случайных величин. Значение f подсчитывается как разность между числом выполненных опытов и числом констант (коэффициентов, средних и т.д.), подсчитанных по результатам тех же опытов.

Уровень значимости α – мера точности ответа. Для инженерных расчетов обычно выбирают $\alpha=0,05$, что соответствует вероятности правильного ответа 0,95 или 95%.

Процесс считается воспроизводимым, т.е. ряд дисперсии однороден, если выполняется неравенство (7.1).

Дисперсия воспроизводимости определяется по формуле

$$S^2_Y = \frac{\sum_{u=1}^n S^2_u}{n}, \quad (7.2)$$

где n – число строк матрицы планирования, т.е. число опытов в плане.

Наряду с дисперсией часто используется еще одна величина – среднее квадратичное отклонение (средняя квадратичная ошибка или стандарт), определяемое как корень квадратный из дисперсии, взятый со знаком плюс:

$$S_Y = \sqrt{S^2_Y} = \sqrt{\frac{\sum_{u=1}^n S^2_u}{n}}. \quad (7.3)$$

Если неравенство (7.1) не выполняется, то необходимо уточнить результаты опытов, имеющих максимальное значение дисперсии.

Необходимо отметить, что по возможности следует избегать случаев, когда число повторений каждого опыта неодинаково. Если почему-либо числа повторений опытов неодинаковы, т.е. дисперсии $S_1^2, S_2^2, S_3^2, \dots$ найдены с различными степенями свободы, однородность дисперсий можно оценить по критерию Бартлета.

В рассматриваемом примере выполняли по два определения величин Y_u и Z_u . Проведенный анализ показал, что процесс воспроизводим с вероятностью 0,95, так как неравенство (7.1) выполняется. При этом дисперсия воспроизводимости по отклику Y составляет $S_Y^2 = 217,25$, а по отклику Z – $S_Z^2 = 0,308$.

Определение коэффициентов уравнения регрессии. По результатам эксперимента определяются коэффициенты математической модели – уравнения регрессии. Все коэффициенты в случае полного факторного эксперимента определяются независимо друг от друга по следующим формулам:

$$b_0 = \frac{\sum_{u=1}^n \bar{Y}_u}{n}; \quad b_{ij} = \frac{\sum_{u=1}^n x_{iu} \bar{Y}_u}{n}; \quad b_i = \frac{\sum_{u=1}^n x_{iu} x_{ju} \bar{Y}_u}{n}, \quad (7.4)$$

где \bar{Y}_u – среднее значение отклика в u -й строке план-матрицы;

x_{iu} – значение i -го фактора в u -й строке план-матрицы.

Для рассматриваемого примера:

$$b_0 = \frac{705 + 1020 + 905 + 1010 + 856 + 1045 + 933 + 1010}{8} = 935,5;$$

$$b_1 = \frac{-705 + 1020 - 905 + 1010 - 856 + 1045 - 933 + 1010}{8} = 85,75.$$

Аналогично определяются другие коэффициенты:

$$b_2 = 29,0; b_3 = 25,5; b_{12} = -40,25; b_{13} = -19,25; b_{23} = -18,5; b_{123} = 11,25;$$

$$b'_0 = 3,825; b'_1 = 2,175; b'_2 = 1,425; b'_3 = 0,55; b'_{12} = 0,675; b'_{13} = 0,25;$$

$$b'_{23} = 0,2; b'_{123} = 0.$$

Оценка значимости коэффициентов уравнений регрессии.

Коэффициент считается значимым, если выполняется неравенство

$$|b_i| \geq \Delta b_i = t_{t(0,05; f_Y)} \sqrt{\frac{S^2_Y}{n}}, \quad (7.5)$$

где $t(0,05; f_Y)$ – значение критерия Стьюдента при 5%-м уровне значимости;

f_Y – число степеней свободы при определении S^2_Y ;

n – число опытов в плане.

Смысл этого неравенства заключается в том, что абсолютная величина коэффициента должна быть в t раз больше, чем ошибка его определения. Средняя квадратичная ошибка определения коэффициентов в условиях полного факторного эксперимента постоянна:

$$S_{bi} = \sqrt{\frac{S^2_Y}{n}}. \quad (7.6)$$

Члены уравнения регрессии с незначимыми коэффициентами отбрасываются, так как данный фактор не влияет или влияет незначительно на параметр оптимизации. Однако на величину коэффициента уравнения регрессии влияет не только роль данного фактора, но также выбранный интервал варьирования. Если интервал варьирования данного фактора увеличить, то его роль может возрасти, что отразится на величине коэффициента уравнения регрессии.

Для рассматриваемого примера значение Δb_i , определяющее доверительный интервал определения коэффициентов в уравнении для прочности сварного соединения, $\Delta b_{Y_i} = 12,04$, а в уравнении для остаточной макропластической деформации сварного соединения $\Delta b_{Z_i} = 0,453$.

Коэффициенты регрессии $b_{123} = 11,25; b'_{13} = 0,25; b'_{23} = 0,2; b'_{123} = 0$ незначимы и их можно не учитывать в математической модели процесса. Таким образом, математическая модель процесса ДСВ сплава ВЖЛ12У со сталью ЭИ961 имеет следующий вид:

$$Y = 935,5 + 85,75x_1 + 29x_2 + 25,5x_3 - 40,25x_1x_2 - 19,25x_1x_3 - 18,5x_2x_3; \quad (7.7)$$

$$Z = 3,825 + 2,175x_1 + 1,425x_2 + 0,55x_3 + 0,675x_1x_2. \quad (7.8)$$

Проверка адекватности математической модели

Адекватность модели проверяется с помощью критерия Фишера:

$$F = \frac{S^2_{ad}}{S^2_Y} \leq F(0,05; f_{ad}; f_Y), \quad S^2_{ad} = \frac{\sum_{u=1}^n (\bar{Y}_u - \hat{Y}_u)^2}{n-1}; \quad (7.9)$$

где \bar{Y} – среднее значение отклика в u -м опыте;

S^2_Y – дисперсия воспроизводимости опытов;

$F(0,05; f_{ad}; f_Y)$ – табличное значение критерия Фишера при 5%-м уровне значимости;

$f_{ad} = n - 1$ – число степеней свободы при определении S^2_{ad} ;

\hat{Y} – расчетное (согласно уравнению регрессии) значение отклика в u -м опыте;

l – число значимых коэффициентов в уравнении регрессии.

Если неравенство (7.9) выполняется, то с вероятностью 0,95 можно считать, что уравнение регрессии адекватно описывает, т.е. достаточно точно аппроксимирует функцию отклика.

Результаты проверки на адекватность математической модели в рассматриваемом примере приведены в табл. 7.5. Как показали расчеты, математическая модель (7.7) неадекватна, т.е. погрешность аппроксимации функции отклика очень велика.

$$S^2_{Y(ad)} = \frac{1200,48}{8-7} = 1200,48 \quad \text{при } f_{ad} = 1; \quad S^2_Y = 217,25;$$

$$F_Y = \frac{1200,48}{217,25} = 5,53 > F_{(0,05;1,8)} = 5,31;$$

$$S^2_{Z(ad)} = \frac{0,82}{8-5} = 0,273 \text{ при } f_{ad} = 3; S^2_Z = 0,308;$$

$$F_Z = \frac{0,273}{0,308} = 0,87 < F_{(0,05;3,8)} = 4,066.$$

Таблица 7.5

Оценка адекватности математической модели

Опыт	x_1	x_2	x_3	\bar{Y}_u	\hat{Y}_u	$(\bar{Y}_u - \hat{Y}_u)^2$	\bar{Z}_u	\hat{Z}_u	$(\bar{Z}_u - \hat{Z}_u)^2$
1	-1	-1	-1	705	717,25	150,06	0,8	0,35	0,2025
2	+1	-1	-1	1020	1007,75	150,06	3,3	3,35	0,0025
3	-1	+1	-1	905	892,75	150,06	1,9	1,85	0,0025
4	+1	+1	-1	1010	1022,25	150,06	7,1	7,55	0,2025
5	-1	-1	+1	856	843,75	150,06	1,0	1,45	0,2025
6	+1	-1	+1	1045	1057,25	150,06	4,5	4,45	0,0025
7	-1	+1	+1	933	945,25	150,06	2,9	2,95	0,0025
8	+1	+1	+1	1010	997,75	150,06	9,1	8,65	0,2025

В связи с этим необходимо увеличить степень полиномов, описывающих зависимости прочности и остаточной макропластической деформации сварного соединения от температуры, давления сжатия и времени сварки. Для решения этой задачи используются специальные планы второго порядка.

Уравнение регрессии (7.8) адекватно описывает зависимость остаточной пластической деформации сварного соединения от температуры, давления сжатия и времени сварки.

Контрольные вопросы

1. Какой эксперимент называется полным факторным экспериментом?
2. Какие обязательные этапы включает в себя метод исследования технологического процесса с помощью планирования эксперимента?
3. В чем заключается этап кодирования факторов?

4. Как определяются интервалы варьирования основных факторов процесса?
5. В чем заключается методика составления план-матрицы эксперимента?
6. В чем заключается процедура рандомизации опытов?
7. Какой вид имеет математическая модель при реализации полного факторного эксперимента?
8. С помощью какого критерия выявляются «грубые» ошибки при реализации эксперимента?
9. В чем заключаются понятия «число степеней свободы» и «уровень значимости»?
10. При каких условиях процесс считается воспроизводимым?
11. В чем заключается методика расчета коэффициентов уравнения регрессии?
12. Какой критерий используется для оценки значимости коэффициентов уравнения регрессии?
13. При каких условиях коэффициент уравнения регрессии является значимым?
14. В чем заключается понятие адекватности математической модели?
15. Какой критерий используется для оценки адекватности математической модели?
16. При каких условиях математическая модель является адекватной?
17. Какие рекомендации могут иметь место в ситуации неадекватности математической модели?

8. ПРИНЯТИЕ РЕШЕНИЙ ПОСЛЕ ПОСТРОЕНИЯ МОДЕЛИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Принимаемые решения зависят от числа факторов, подробности плана, цели исследования (достижение оптимума, построение интерполяционной формулы) и т.д. Основные варианты ситуаций, возникающих после построения модели, заключаются в различных комбинациях следующих полученных результатов: адекватность или неадекватность модели, значимость или незначимость коэффициентов регрессии в модели, место расположения оптимума в факторном пространстве.

Предположим, что полученная модель линейна и адекватна. Здесь возможны три варианта ситуации: все коэффициенты регрессии значимы; часть коэффициентов регрессии значима, а часть их незначима; все коэффициенты регрессии незначимы.

В каждом варианте оптимум может быть близко, далеко или о его положении нет информации (неопределенная ситуация). Если область оптимума близка, возможны три решения: окончание исследования, переход к планам второго порядка и движение по градиенту.

Первый вариант – все коэффициенты регрессии значимы. Переход к планированию второго порядка дает возможность получить математическое описание области оптимума и найти экстремум. Движение по градиенту используется при малой ошибке опыта, поскольку на фоне большой ошибки трудно установить приращение параметра оптимизации. Решение при неопределенной ситуации или удаленной области оптимума одно и то же: движение по градиенту.

Второй вариант – часть коэффициентов регрессии значима, часть незначима. Движение по градиенту наиболее эффективно в случае значимости коэффициентов. Поэтому выбираются решения, реализация которых приводит к получению значимых коэффициентов. Важно выдвинуть гипотезы, объясняющие незначимость эффектов. Это может быть и неудачный выбор

интервалов варьирования, и включение факторов, действительно не оказывающих влияния на параметр оптимизации, и большая ошибка опыта, и т.д. Если, например, выдвинута первая гипотеза, то возможно такое решение: расширение интервалов варьирования по незначимым факторам и постановка новой серии опытов. Изменение интервалов варьирования иногда сочетают с переносом центра эксперимента в точку, соответствующую условиям наилучшего опыта. Невлияющие факторы стабилизируются и исключаются из дальнейшего рассмотрения.

Другие возможные решения для получения значимых коэффициентов: увеличение числа параллельных опытов и доработка плана. Увеличение числа параллельных опытов приводит к уменьшению дисперсии воспроизводимости и соответственно дисперсии коэффициентов регрессии. Опыты могут быть повторены либо во всех точках плана, либо в некоторых.

Доработка плана осуществляется несколькими способами: у исходной реплики можно изменить знаки на обратные (в этом случае основные эффекты оказываются не смешанными с парными эффектами взаимодействия); перейти к полному факторному эксперименту; перейти к реплике меньшей дробности; перейти к плану второго порядка (если область оптимума близка). Реализация любого из этих решений требует значительных экспериментальных усилий. Поэтому нужно начинать с движения только по значимым факторам.

Наконец, если область оптимума близка, то возможно принятие таких же решений, как и в случае значимости всех коэффициентов регрессии.

Третий вариант – линейная модель адекватна, все коэффициенты регрессии незначимы (кроме b_0). Чаще всего это происходит вследствие большой ошибки эксперимента или узких интервалов варьирования. Возможные решения состоят, прежде всего, в направлении увеличения точности эксперимента и расширения интервалов варьирования. Увеличение точности может достигаться двумя путями: улучшением методики проведения опытов или постановкой параллельных опытов.

В задаче построения интерполяционной формулы цель исследования считают достигнутой, если получена адекватная модель. Обобщающая блок-схема принятия решения в задаче определения оптимальных условий при адекватной линейной модели показана на рис. 8.1. Пунктирными линиями обведены ситуации, сплошными линиями – принимаемые решения.

Если линейная модель неадекватна, значит, не удастся аппроксимировать поверхность отклика плоскостью. Формальные признаки неадекватности линейной модели следующие:

- расчетная величина F -критерия превышает табличное значение;
- значимость хотя бы одного из эффектов взаимодействия;
- значимость суммы коэффициентов регрессии при квадратичных членах.



Рис. 8.1. Принятие решений в задаче определения оптимальных условий адекватности линейной модели:

— принимаемые решения; - - - ситуации

Оценкой значимости этой суммы служит разность между b_0 и полученным в эксперименте значением отклика в центре плана y_0 . Если разность превосходит ошибку опыта, то гипотеза о незначимости коэффициентов при квадратичных членах не может быть принята. Однако надо

учесть, что сумма может быть незначима и при значимых квадратичных эффектах, если они имеют разные знаки.

Для неадекватной модели мы не будем делать различия между случаями значимых и незначимых линейных коэффициентов регрессии, поскольку решения для них обычно совпадают.

Решения, принимаемые для получения адекватной модели, состоят в следующем: изменение интервалов варьирования факторов, перенос центра плана, достройка плана. Наиболее распространенный прием – изменение интервалов варьирования, однако при этом потребуются постановка новой серии опытов. Еще одно решение: включение в модель эффектов взаимодействия и движение с помощью неполного полинома второго порядка. Этот прием связан с получением и анализом уравнений второго порядка. Направление градиента будет меняться от точки к точке. Если область оптимума близка, то возможны варианты окончания исследования и перехода к построению плана второго порядка.

На рис. 8.2 приведена блок-схема принятия решений в задаче оптимизации для случая, когда линейная модель неадекватна.

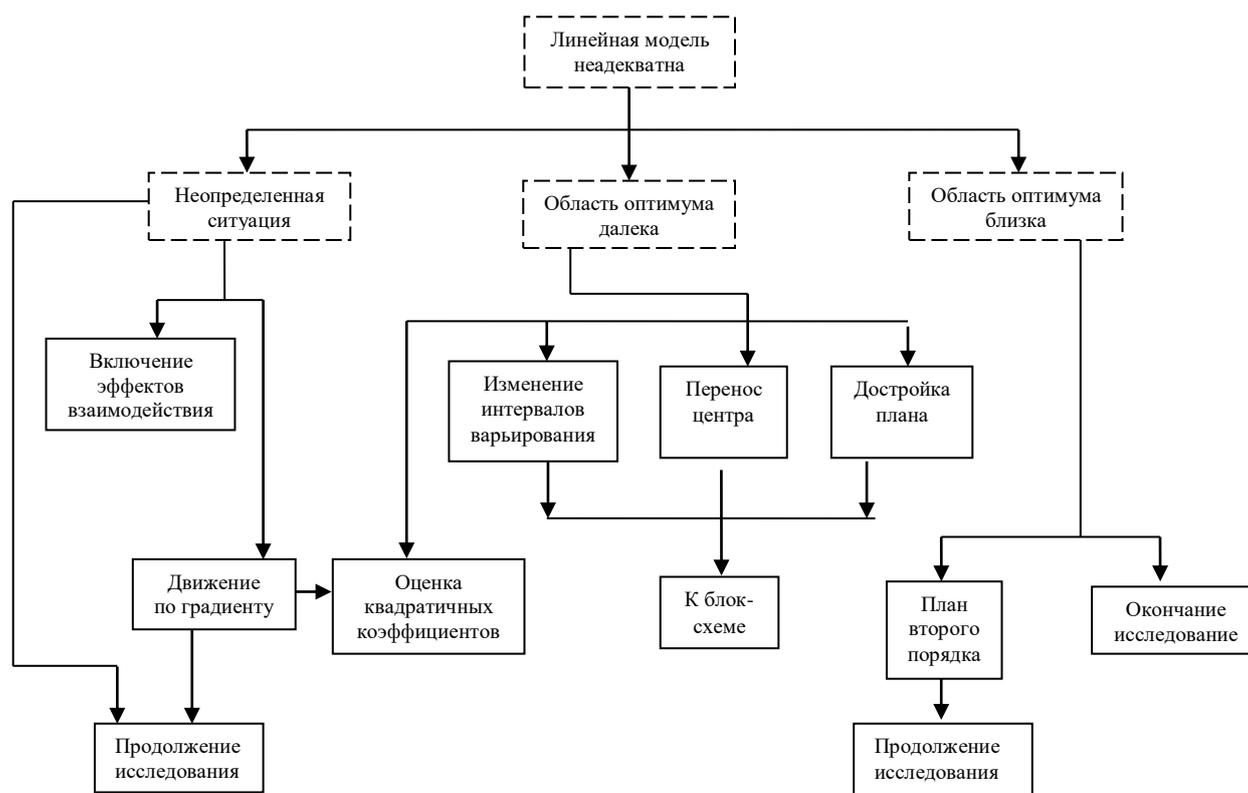


Рис. 8.2. Принятие решений в задаче определения оптимальных условий при неадекватности линейной модели

Особый случай возникает при использовании насыщенных планов. При значимости всех коэффициентов регрессии ничего нельзя сказать об адекватности или неадекватности модели. Движение по градиенту в такой ситуации показывает правильность предположения, что коэффициенты регрессии являются оценками для линейных эффектов.

Построение интерполяционной формулы: линейная модель неадекватна. Первое, что следует сделать при решении этой задачи, – включить в уравнение эффекты взаимодействия. Такое решение может быть реализовано, если был применен ненасыщенный план. После введения эффектов взаимодействия может не хватать степеней свободы и потребуются реализация еще двух-трех опытов внутри области эксперимента для проверки гипотезы адекватности. Чаще всего для построения интерполяционной формулы необходимо проведение новых опытов. Следует провести достройку плана. Используются те же приемы, что и при устранении незначимости коэффициентов регрессии: перемена знаков у исходной реплики, достройка до полного факторного эксперимента или до дробной реплики, в которой исключены ранее смешанные эффекты, а также достройка до плана второго порядка. Часто помогает еще один прием – преобразование зависимых и независимых переменных, при котором натуральные значения факторов заменяются логарифмическими или показательными функциями [3, 5, 9, 13].

Если все перечисленные приемы не помогли и адекватную модель получить не удалось, приходится разбивать область эксперимента на несколько подобластей и описывать отдельно каждую из них. Для этого необходимо уменьшить интервалы варьирования факторов.

На рис. 8.3 приведена блок-схема принятия решений в задаче построения интерполяционной формулы для случая, когда линейная модель неадекватна. Если линейная модель адекватна, то задача решена.

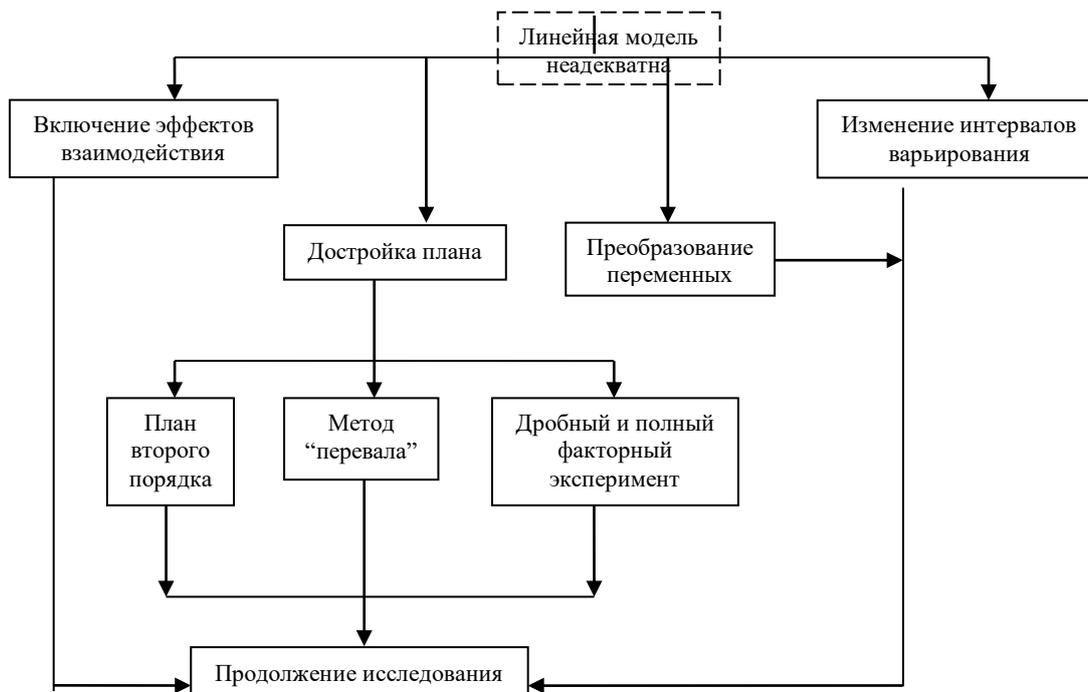


Рис. 8.3. Принятие решений в задаче построения интерполяционной формулы при неадекватности линейной модели

Контрольные вопросы

1. Какие методы применяются для интерпретации результатов факторного эксперимента?
2. Назовите основные этапы интерпретации модели.
3. Что означает термин «значимость коэффициента модели»?
4. В чем заключается важность априорной информации при планировании эксперимента?
5. Какие рекомендации могут иметь место в случае неадекватности математической модели?

9. ПЛАНЫ ДРОБНОГО ФАКТОРНОГО ЭКСПЕРИМЕНТА

При многофакторном эксперименте, особенно когда число факторов больше шести ($n > 6$), число опытов планов ПФЭ 2^n ($N = 2^n$) становится чрезмерным. Если нам не требуется определение всех коэффициентов неполного квадратичного полинома, то переходят к дробному факторному эксперименту (ДФЭ) – части полного факторного эксперимента. Так, например, если требуется определить лишь коэффициенты при самих факторах

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n,$$

то план ПФЭ 2^n дает избыточную информацию. Так, при $n = 6$ в этом случае требуется определить $n+1 = 7$ коэффициентов, тогда как по плану ПФЭ необходимо провести $N = 2^6 = 64$ опыта.

Хотя эта избыточная информация не является бесполезной, она позволяет более точно определить коэффициенты, но все же часто используют планы ДФЭ 2^{n-k} , где k – показатель дробности плана ПФЭ. При $k = 1$ число опытов в плане ДФЭ в два раза меньше, чем в плане ПФЭ, поэтому такие планы называют полуреплика плана ПФЭ. Так при $k=1$ для плана ДФЭ 2^{6-1} $N = 2^{6-1} = 32$, при $k=2$ для плана ДФЭ 2^{6-2} $N = 2^{6-2} = 16$ и такой план называют четвертьрепликой, при $k=3$ для плана ДФЭ 2^{6-3} $N = 2^{6-3} = 8$. При выборе дробности плана k необходимо учитывать, что число опытов должно быть больше числа членов уравнения. В рассматриваемом случае величина k должна быть такой, чтобы удовлетворялось условие

$$n+1 \leq 2^{n-k}.$$

План ДФЭ строится так же, как и для плана ПФЭ, но с меньшим числом факторов. Оставшиеся факторы варьируются не произвольно, а так, чтобы сохранялась ортогональность плана. Это обеспечивается, если оставшиеся факторы варьируются по выбранному генерирующему соотношению, например как произведение каких-либо факторов из первой группы. Но это приводит к тому, что в матрице X будут существовать одинаковые столбцы. Следовательно, мы не сможем найти в чистом виде все коэффициенты

неполного квадратичного полинома, а лишь определим совместную величину коэффициентов для одинаковых столбцов.

Рассмотрим построение плана ДФЭ 2^{3-1} . Здесь $n = 3$, $K = 1$, $N = 2^{3-1} = 4$. Первые два фактора варьируем так же, как и ранее для плана ПФЭ 2^2 , а для третьего фактора выбираем генерирующее соотношение в виде $x_3 = x_1 \cdot x_2$.

Для неполного квадратичного полинома

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$$

количество столбцов плана составляет восемь.

i	0	1	2	3	4	5	6	7
U	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	+1	-1	-1	+1	+1	-1	-1	+1
2	+1	+1	-1	-1	-1	-1	+1	+1
3	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	+1	+1	+1	+1	+1	+1

План является ортогональным, но в нем оказались четыре пары одинаковых столбцов. Поэтому можно определить только четыре коэффициента, отражающие совместные влияния двух одинаковых столбцов:

$$b_0 + b_{123} = \frac{\sum_{U=1}^N x_{0U} Y_U}{N}.$$

Суммарные значения коэффициентов $b_1 + b_{23}$; $b_2 + b_{13}$; $b_3 + b_{12}$ определяются аналогично. Это следствие того, что мы пытаемся определить полное количество коэффициентов – 8 по недостаточному числу опытов – 4. Однако, если заранее известно, что некоторые из членов уравнения равны нулю (пренебрежимо малы) или имеется априорная информация о величинах некоторых коэффициентов, то полученные коэффициенты могут быть вычленены. Так, если $b_{123} = 0$, то

$$b_0 = \frac{\sum_{U=1}^N x_{0U} Y_U}{N}.$$

Если можно допустить, что коэффициенты из их смешанной оценки сопоставимы, то для рассмотренного плана

$$b_0 = b_{123} = \frac{1}{2} \frac{\sum_{\nu=1}^N x_{0\nu} Y_{\nu}}{N}.$$

Графическое изображение планов ПФЭ 2^3 и ДФЭ 2^{3-1} в факторном пространстве (для трех факторов – трехмерное пространство) представлено на рис. 9.1. План ПФЭ 2^3 представлен кубом с восемью узлами (точками плана), а возможные планы ДФЭ 2^{3-1} – проекциями этого куба на три плоскости. То есть из восьми узлов выбираются четыре (рис. 9.1, а). Из куба можно также выбрать четыре точки из восьми, не лежащие в одной плоскости, и сформировать план ДФЭ 2^{3-1} (рис. 9.1, б).

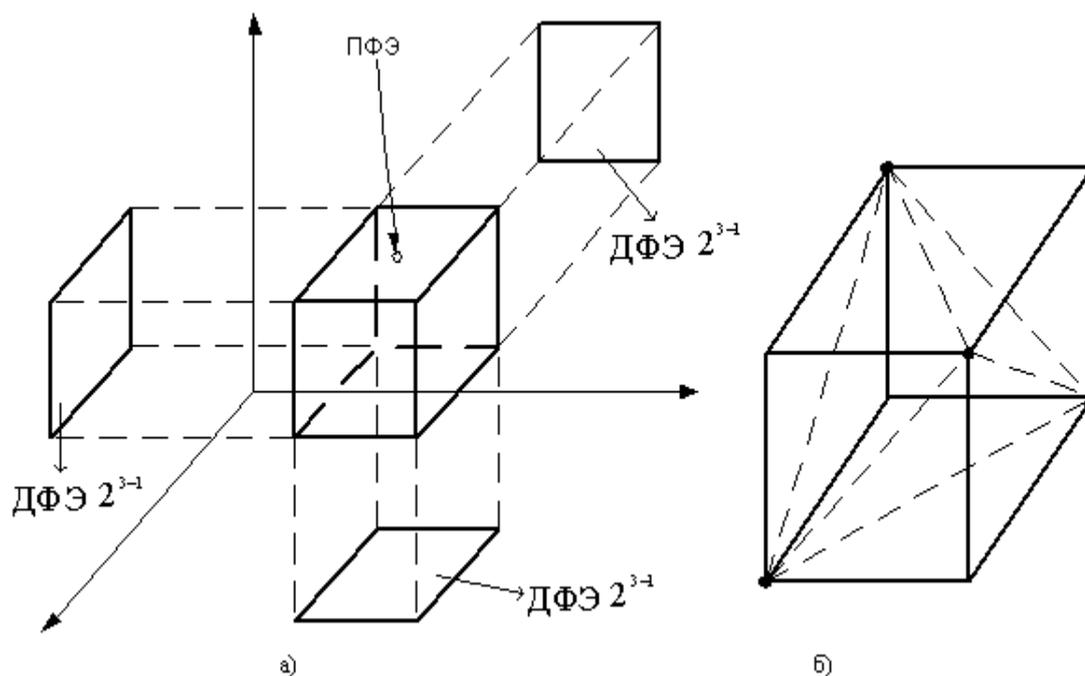


Рис. 9.1. Графическое изображение планов ПФЭ 2^3 и ДФЭ 2^{3-1} в факторном пространстве

Планы ДФЭ, как и планы ПФЭ, являются ротатабельными. Планы ДФЭ могут быть как насыщенными, так и ненасыщенными.

Достоинство планов ДФЭ заключается и в том, что если построенный на его основе неполный полином не удовлетворяет требованиям по точности, то

план ДФЭ легко достраивается до плана ПФЭ без потери информации прежних опытах, с формированием более точного полинома.

Пример. Построить план ДФЭ.

Построить план ДФЭ 2^{4-1} и определить полином

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4.$$

Число факторов – 4. Нужно найти 8 коэффициентов полинома. Выбираем 8 из 16 опытов плана ПФЭ 2^4 таким образом, чтобы были определены независимые коэффициенты при самих факторах, смешанные коэффициенты при парных сочетаниях факторов и в пренебрежении тройными и четверным сочетаниями факторов и при этом сохранялась ортогональность плана.

U		x_1	x_2	x_3	x_4
ПФЭ 2^4	ДФЭ 2^{4-1}				
1	1	-1	-1	-1	-1
2	-	+1	-1	-1	-1
3	-	-1	+1	-1	-1
4	4	+1	+1	-1	-1
5	-	-1	-1	+1	-1
6	6	+1	-1	+1	-1
7	7	-1	+1	+1	-1
8	-	+1	+1	+1	-1
9	-	-1	-1	-1	+1
10	2	+1	-1	-1	+1
11	3	-1	+1	-1	+1
12	-	+1	+1	-1	+1
13	5	-1	-1	+1	+1
14	-	+1	-1	+1	+1
15	-	-1	+1	+1	+1
16	8	+1	+1	+1	+1

Такой выбор позволяет сформировать план ДФЭ 2^{4-1} как и план ПФЭ 2^3 , но с $x_4 = x_1x_2x_3$. План ДФЭ 2^{4-1} представляется в виде

U	x_0	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4	Y	\hat{Y}
1	+1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	10	10
2	+1	+1	-1	-1	+1	-1	-1	+1	+1	-1	-1	8	8
3	+1	-1	+1	-1	+1	-1	+1	-1	-1	+1	-1	8	8
4	+1	+1	+1	-1	-1	+1	-1	-1	-1	-1	+1	7	6,8
5	+1	-1	-1	+1	+1	+1	-1	-1	-1	-1	+1	9	9,2
6	+1	+1	-1	+1	-1	-1	+1	-1	-1	+1	-1	8	8
7	+1	-1	+1	+1	-1	-1	-1	+1	+1	-1	-1	8	8
8	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	6,5	6,68



Значения коэффициентов полинома составляют:

$$b_{123} = b_{124} = b_{134} = b_{234} = b_{1234} = 0;$$

$$b_0 = \frac{10 + 8 + 8 + 7 + 9 + 8 + 8 + 6,5}{8} = 8,06;$$

$$b_1 = \frac{-10 + 8 - 8 + 7 - 9 + 8 - 8 + 6,5}{8} = -0,69;$$

$$b_2 = -0,69; \quad b_3 = -0,19; \quad b_4 = -0,19;$$

$$(b_{12} + b_{34}) = \frac{10 - 8 - 8 + 7 + 9 - 8 - 8 + 6,5}{8} = 0,06;$$

$$b_{13} + b_{24} = 0,06, \quad b_{14} + b_{23} = 0,06.$$

Если принять, что

$$b_{12} \approx b_{34} = \frac{1}{2}(b_{12} + b_{34}) = 0,03;$$

$$b_{13} \approx b_{24} = \frac{1}{2}(b_{13} + b_{24}) = 0,03;$$

$$b_{14} \approx b_{23} = \frac{1}{2}(b_{14} + b_{23}) = 0,03,$$

то полином имеет вид

$$\hat{Y} = 8,06 - 0,69x_1 - 0,69x_2 - 0,19x_3 - 0,19x_4 + 0,03x_1x_2 + \\ + 0,03x_1x_3 + 0,03x_1x_4 + 0,03x_2x_3 + 0,03x_2x_4 + 0,03x_3x_4.$$

Значения полинома в точках плана приведены в последнем столбце плана ДФЭ 2^{4-1} . В нашем случае точность его достаточно высокая.

Контрольные вопросы

1. В чем заключается идея дробного факторного эксперимента?
2. Основные преимущества использования дробного факторного эксперимента.
3. Какие недостатки имеет дробный факторный эксперимент?

10. ОПТИМИЗАЦИЯ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

10.1. Понятие о задаче условной оптимизации. Классификация задач оптимизации

Ранее мы уже описали общую постановку задачи оптимизации и изучили простейший класс оптимизационных задач без ограничений. В данной лекции мы продолжим разговор о задачах оптимизации как математической основе теории управления технологическими процессами.

Задачи оптимизации находят эффективное применение во всех направлениях инженерной деятельности, и в первую очередь в следующих четырех ее областях.

Проектирование систем и их составных частей. Сфера применения оптимизационных методов в инженерном проектировании достаточно широка: от проектирования отдельных структурных элементов технических систем до проектирования узлов оборудования и составления предварительных проектов промышленных предприятий в целом.

Планирование и анализ функционирования существующих систем. Методы оптимизации в производственном планировании ориентированы главным образом на составление программ производства нескольких видов продукции на отдельном предприятии, а также на координирование производственных планов предприятий, которые связаны хозяйственными отношениями. Задачи анализа функционирования систем обычно возникают в тех случаях, когда требуется адаптировать существующую производственную систему к новым условиям функционирования, отличным от тех условий, которые были предусмотрены проектом этой системы.

Инженерный анализ и обработка информации. Эта область применения оптимизационных методов в инженерной практике связана с задачами инженерного анализа, в частности с задачами нелинейного регрессионного анализа. Среди наиболее общих проблем, возникающих в процессе разработки инженерных моделей, можно выделить проблему определения параметров

полуэмпирической модели на основе заданного множества экспериментальных данных. Такого рода задачи приводятся к некоторому виду оптимизационных задач.

Управление динамическими системами. В этой области находит широкое применение наиболее сложный раздел теории оптимизации – математическая теория оптимального управления. Динамические системы могут быть описаны при помощи дифференциальных, интегральных или иных уравнений.

Для любой практической оптимизационной задачи можно выделить следующие основные этапы её реализации:

1. Моделирование рассматриваемой физической ситуации с целью получения математической функции, которую необходимо оптимизировать, а также определение ограничений для параметров.
2. Проверка задачи на существование и единственность решения.
3. Выбор подходящей математической процедуры (метода) для осуществления оптимизации.
4. Реализация выбранной процедуры на практике, в частности на ЭВМ.
5. Анализ математического результата и интерпретация его в терминах физического содержания задачи.

Вернемся к общей постановке задачи оптимизации. Как уже отмечалось в предыдущей лекции, с математической точки зрения различия между задачами минимизации и максимизации несущественны. Поэтому мы всегда можем ставить задачу оптимизации как задачу минимизации (10.1):

$$f(x) \rightarrow \min, x \in X. \quad (10.1)$$

Так же как и в предыдущей лекции, функцию $f(x)$ в (10.1), т.е. функцию, которую мы минимизируем, будем называть *целевой функцией*, множество X в (10.1), на котором мы минимизируем $f(x)$, – *допустимым множеством* задачи (10.1), любой элемент $x \in X$ – *допустимой точкой* задачи (10.1). Допустимую точку $x^* \in X$, в которой целевая функция $f(x)$ достигает своего минимума, будем называть *решением* задачи (10.1).

В задачах условной оптимизации допустимая точка x может представлять собой некий набор параметров $x = (x_1, \dots, x_n)$. Значения этих параметров подчиняются некоторым ограничениям. Если, например, x_i выражает количество производимого продукта i -го вида (сплав соответствующей марки и пр.) для каждого $i = 1, \dots, n$, то при этом будет существовать ограничение на производственную мощность (например, $0 \leq x_i \leq a_i, i = 1, \dots, n$).

Необходимо подчеркнуть, что само понятие точки минимума (решения задачи (10.1)) неоднозначно и требует уточнения.

Определение. Точка $x^* \in X$ называется точкой *глобального минимума* функции $f(x)$ на множестве X или *глобальным решением* задачи (10.2), если

$$f(x^*) \leq f(x) \text{ при всех } x \in X. \quad (10.2)$$

Точка $x^* \in X$ называется точкой *локального минимума* f на X или *локальным решением* задачи (10.2), если существует такое подмножество

$$U_\delta(x^*) = \{x / x \in X, \|x - x^*\| \leq \delta\}, \text{ что}$$

$$f(x^*) \leq f(x) \text{ при всех } x \in U_\delta(x^*). \quad (10.3)$$

Здесь и далее

$$\|x - x^*\| = \left[\sum_{i=1}^n (x_i - x_i^*)^2 \right]^{0.5}.$$

Очевидно, что глобальное решение является и локальным; обратное неверно.

Определение. Если неравенство в (10.2) или (10.3) выполняется как строгое при $x \neq x^*$, т. е. $f(x) < f(x^*), x \in X$ (или $x \in U_\delta(x^*)$), то говорят, что x^* – точка *строгого минимума* (*строгое решение*).

Рассмотрим пример, иллюстрирующий введённые понятия.

П р и м е р. Пусть $f(x) = x(x^2 - 3)$. На отрезке $X = [-1,5; 2]$ функция $f(x)$ достигает строгого локального минимума в точке $x = -1,5$ и строгого глобального минимума в точке $x = 1$. На отрезке $X = [-2, 2]$ она достигает глобального минимума в двух точках: $x = -2, x = 1$ ($f(-2) = f(1) = -2$).

При изучении задач оптимизации в первую очередь возникает вопрос о *существовании решения*. Условия, гарантирующие разрешимость задачи (10.2), содержатся в следующей теореме из курса высшей математики.

Теорема (Вейерштрасса). Пусть X – ограниченное замкнутое множество в \mathbf{R}^n , $f(x)$ – непрерывная функция на X . Тогда существуют точки глобального минимума и максимума функции f на X (глобальное решение задачи (10.2)).

Эта теорема отвечает на вопрос о существовании решения, но не даёт конструктивного алгоритма нахождения этого решения. Следует также учесть, что задача оптимизации может иметь несколько решений. Вопрос о единственности решения рассматривается отдельно для каждого класса задач.

Классификацию задач оптимизации можно проводить по нескольким признакам в зависимости от вида исходных данных. Выделим наиболее важные из них.

Задачи безусловной оптимизации – это задачи (10.2), в которых допустимым множеством X является все пространство \mathbf{R}^n .

Задачами математического программирования называются задачи (10.2), если их допустимое множество задается системой конечного числа неравенств и уравнений, т. е. $X = \{x \in P \mid g_i(x) \leq 0, i = 1, \dots, k; h_i(x) = 0, i = k + 1, \dots, m\}$.

Среди них важнейшими являются следующие классы задач:

- ◆ задачи линейного программирования;
- ◆ задачи нелинейного программирования (задачи квадратичного программирования, задачи выпуклого программирования и др.).

Задачами дискретного программирования называются задачи (10.2), если какие-либо из координат x_1, \dots, x_n пробегают дискретные множества на числовой оси, когда x пробегает X .

Задачи оптимального управления относятся к классу бесконечномерных задач оптимизации, где в качестве допустимых множеств, по которым ведется минимизация, выступают множества функций.

Понятие о численных методах оптимизации

Любой численный метод (алгоритм) решения задачи оптимизации основан на точном или приближенном вычислении ее *характеристик* (значений целевой функции и функций, задающих границу допустимого множества, а также их производных). На основе полученной информации строится приближение к решению задачи – искомой точке минимума x^* или, если такая точка не единственна, к множеству точек минимума. Иногда, если это требуется, строится приближение к минимальному значению целевой функции f^* .

Алгоритмы, использующие лишь информацию о значениях минимизируемой функции, называются алгоритмами *нулевого порядка*; алгоритмы, использующие также информацию о значениях первых производных, – алгоритмами *первого порядка*; алгоритмы использующие, кроме того, информацию о вторых производных, – алгоритмами *второго порядка*.

Работа алгоритма состоит из двух этапов. На *первом этапе* вычисляются предусмотренные алгоритмом характеристики задачи. На *втором этапе* по полученной информации строится приближение к решению. Обычно для задания алгоритма достаточно указать способ выбора точек приближения

$x^k = (x_1^k, \dots, x_n^k) \in X$, $k = 1, 2, \dots$ (конечно, при условии, что уже решен вопрос о том, какие именно характеристики задачи следует вычислять).

Определение. Выбор точек приближения называется *поиском* точек.

Если все точки выбираются одновременно до начала вычислений, то алгоритм минимизации называется *пассивным*. Однако для решения большинства задач точки приближения выбираются поочередно, т.е. точка $x^{(k+1)}$ выбирается тогда, когда уже выбраны точки предыдущих вычислений $x^{(1)}$, $x^{(2)}$, ..., $x^{(k)}$ и в каждой из них произведены предусмотренные алгоритмом вычисления. Такие алгоритмы называются *последовательными*.

Определение. В последовательных алгоритмах вычисление характеристик задачи в точке $x^{(k)}$ и поиск точки $x^{(k+1)}$ вместе составляют *шаг метода*, или, что то же самое, *итерацию метода*.

Методы оптимизации можно условно разделить на две основные группы: конечношаговые и бесконечношаговые. *Конечношаговыми*, или *конечными*, называются методы, гарантирующие отыскание решения задачи за конечное число шагов. *Бесконечношаговыми* называются методы, для которых достижение решения гарантируется лишь в пределе.

Важной характеристикой бесконечношаговых методов оптимизации является сходимость

Определение. Будем говорить, что метод *сходится*, если последовательность $x^{(k)} \rightarrow x^*$ при $k \rightarrow \infty$, где x^* – решение задачи (10.2). Если $f(x^{(k)}) \rightarrow f(x^*)$ при $k \rightarrow \infty$, то последовательность $x^{(k)}$ называют *минимизирующей*.

Заметим, что минимизирующая последовательность может и не сходиться к точке минимума. Так для функции, график которой изображен на рис. 10.1, минимизирующая последовательность $x^{(k)} = k$ не сходится к точке минимума $x^* = 0$.

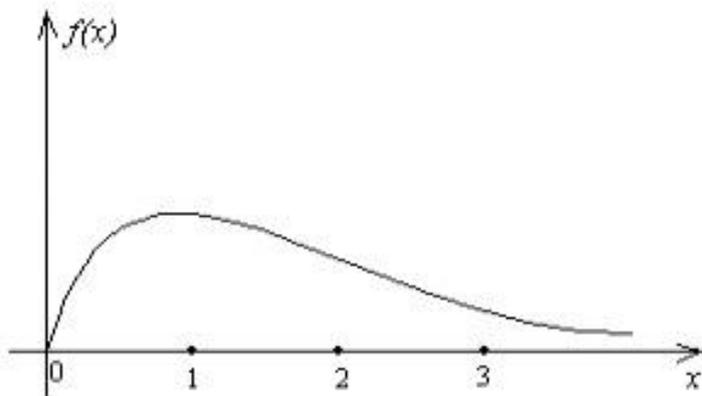


Рис. 10.1. Пример функции, для которой существует расходящаяся минимизирующая последовательность

В случае, когда точка минимума x^* не единственна, под сходимостью метода понимается следующее: для каждой точки x^* минимума функции f можно построить данным методом такую последовательность $x^{(k)} \in X$, что $x^{(k)} \rightarrow x^*$ при $k \rightarrow \infty$.

Установление факта сходимости дает существенную информацию о выбранном методе минимизации. Прежде всего, требования, которые приходится налагать для обеспечения сходимости на минимизируемую функцию, показывают область применимости метода. Часто условия сходимости включают в себя в явном виде требования к начальному приближению.

В то же время реальный процесс оптимизации не может быть бесконечношаговым. Кроме того, в ряде случаев условия сходимости труднопроверяемы. Поэтому при выборе подходящего метода решения реальных задач приходится во многом руководствоваться здравым смыслом, опытом, интуицией, а также результатами численных экспериментов. Необходимо также учитывать погрешность исходных данных.

Условия оптимальности в общей задаче оптимизации

При изучении любого типа задач оптимизации важное место занимает вопрос об *условиях оптимальности*, или, как еще говорят, условиях экстремума. Различают *необходимые* условия оптимальности, т.е. условия, которым должна удовлетворять точка, являющаяся решением задачи, и *достаточные* условия оптимальности, т.е. условия, из которых следует, что данная точка является решением задачи. Интерес к условиям оптимальности объясняется тем, что они, во-первых, составляют основу качественных методов теории оптимизации, т.е. методов, направленных на изучение свойств задач оптимизации; во-вторых, используются при построении и обосновании численных методов решения этих задач; в третьих, позволяют в простых случаях явно решить задачу. Примером таких условий являются необходимые и достаточные условия локального экстремума, рассмотренные в предыдущей лекции.

В общем случае суть всех критериев оптимальности для задачи (10.2) заключается в том, что из точки x^* , являющейся локальным решением, нельзя осуществить сколь угодно малый сдвиг в каком бы то ни было направлении,

приводящий к уменьшению значения целевой функции, и остаться при этом в пределах допустимого множества.

10.2. Крутое восхождение по поверхности отклика. Понятие градиента

Любую совокупность вещественных чисел (v_1, v_2, \dots, v_k) , взятых в определенном порядке, можно рассматривать как точку или вектор с теми же координатами в пространстве k измерений (k -мерном пространстве). Запись вида $\mathbf{v} = (v_1, v_2, \dots, v_k)$ обозначает точку или вектор \mathbf{v} с указанными в скобках координатами. Если для k -мерных векторов \mathbf{v} и \mathbf{w} справедливы основные алгебраические операции:

сложение и вычитание

$$\mathbf{v} \pm \mathbf{w} = (v_1 \pm w_1, v_2 \pm w_2, \dots, v_k \pm w_k),$$

умножение на действительное число u

$$u \times \mathbf{v} = (u \times v_1, u \times v_2, \dots, u \times v_k),$$

скалярное произведение

$$\mathbf{v} \times \mathbf{w} = (v_1 \times w_1, v_2 \times w_2, \dots, v_k \times w_k),$$

то совокупность всех таких векторов называют k -мерным евклидовым пространством и обозначают E_k .

Длиной вектора \mathbf{v} называют число, определяемое по формуле

$$|\mathbf{v}| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \dots + v_k^2}. \quad (10.4)$$

Длину вектора можно вычислить только тогда, когда компоненты вектора представлены в одной шкале измерений или они являются безразмерными величинами, полученными, например, в результате преобразования – кодированные переменные безразмерны.

Если произведение $\mathbf{v} \times \mathbf{w} = 0$ при $|\mathbf{v}| \neq 0$ и $|\mathbf{w}| \neq 0$, то векторы \mathbf{v} и \mathbf{w} являются ортогональными.

Единичным называют вектор, определяемый по формуле

$$\mathbf{t} = (t_1, t_2, \dots, t_k) = \left(\frac{v_1}{|\mathbf{v}|}, \frac{v_2}{|\mathbf{v}|}, \dots, \frac{v_k}{|\mathbf{v}|} \right). \quad (10.5)$$

Пусть в E_k заданы некоторая точка $V = (v_1, v_2, \dots, v_k)$, единичный вектор \mathbf{t} и непрерывно дифференцируемая по всем аргументам функция $f(V) = f(v_1, v_2, \dots, v_k)$. Производной в точке V от функции $f(V)$ по направлению луча, определяемому вектором \mathbf{t} , называется предел

$$\frac{\partial f(V)}{\partial \mathbf{t}} = \lim_{\lambda \rightarrow 0} \frac{f(v_1 + \lambda t_1, v_2 + \lambda t_2, \dots, v_k + \lambda t_k) - f(v_1, v_2, \dots, v_k)}{\lambda}$$

или

$$\frac{\partial f(V)}{\partial \mathbf{t}} = \left[\frac{\partial f(V)}{\partial v_1} t_1, \frac{\partial f(V)}{\partial v_2} t_2, \dots, \frac{\partial f(V)}{\partial v_k} t_k \right].$$

Градиентом функции $f(V)$ называют вектор $\nabla f(V)$ с координатами, равными частным производным по соответствующим аргументам

$$\nabla f(V) = \left(\frac{\partial f(V)}{\partial v_1}, \frac{\partial f(V)}{\partial v_2}, \dots, \frac{\partial f(V)}{\partial v_k} \right). \quad (10.6)$$

Градиент указывает направление наибольшего возрастания функции. Противоположное направление $-\nabla f(V)$ называется антиградиентом, оно показывает направление наискорейшего убывания функции. В точке экстремума V^* градиент равен нулю $\nabla f(V^*) = 0$. Если аналитически производные определить невозможно, их вычисляют приближенно $\partial f(V) / \partial v_i \approx \Delta f(V) / \Delta v_i$, где $\Delta f(V)$ – приращение функции $f(V)$ при изменении аргумента на величину Δv_i . Двигаясь по градиенту (антиградиенту), можно достичь максимума (минимума) функции. В этом и состоит сущность градиентного метода оптимизации.

10.3. Способы градиентной оптимизации

Существует несколько модификаций метода градиентной оптимизации применительно к дискретным вычислениям.

Если подъем происходит поочередно по каждой отдельной координате v_1, v_2, \dots, v_k , то такой метод называют *покоординатным подъемом*, или методом Гаусса – Зейделя. Движение осуществляется из некоторой точки по координате v_1 до тех пор, пока не станет равной нулю соответствующая производная $\partial f(V) / \partial v_1 = 0$. Все остальные координаты (аргументы функции)

сохраняют постоянное значение. После этого подъем начинается по другой координате. Порядок перебора координат не играет принципиальной роли, а влияет только на скорость поиска, поэтому обычно начинают с v_1 , затем с v_2 и т.д. После того как будет произведен подъем по всем координатам, начинают повторно с v_1 . Процесс заканчивается, когда все частные производные будут равны нулю (будут меньше порога чувствительности).

Метод наискорейшего подъема предполагает определение градиента в исходной точке, далее подъем в этом направлении осуществляется до тех пор, пока производная $df(V) / dV$ в этом направлении не обратится в нуль. После этого снова определяют градиент и осуществляют по нему подъем до нулевого значения производной и т.д. Модификация этого метода предусматривает вычисление градиента в каждой новой точке траектории перемещения.

Все сказанное о сущности методов поиска максимума функции легко транспонируется и для поиска минимума. Рассмотренные выше методы предполагают возможность движения по любому выбранному направлению, т.е. ограничений на область допустимых значений аргументов нет. В качестве начальной точки может быть взята любая точка пространства E_k . Равенство $\nabla f(V^*) = 0$ является необходимым, но не достаточным условием экстремума функции в точке V^* , да и точек V^* может быть несколько. Поэтому требуются дополнительные исследования для установления, какая из точек действительно является оптимумом (а не точкой перегиба) и какая из них является глобальной.

Одна из основных проблем применения градиентного метода поиска заключается в выборе величины каждого дискретного шага. Шаги могут быть постоянными или переменными. Второй вариант в реализации алгоритма более сложный, но обычно требует меньшего количества итераций.

Поиск максимума функции включает следующие этапы.

1. Определение аналитических соотношений для вычисления градиента функции $\nabla f(V)$, длины вектора градиента $|\nabla f(V)|$ и единичного вектора $t(V)$, используя соответственно формулы (10.4), (10.5) и (10.6).

2. Выбор исходной точки V_n при $n = 0$ (начальных значений аргументов функции).

3. Вычисление координат единичного вектора $t(V_n)$ по формуле, полученной на шаге 1, и определение координат новой точки при движении по направлению единичного вектора.

4. Выбор шага a изменения координат текущей точки V_n . Осуществляется из условия предельного увеличения функции $f[V_n + at(V_n)]$ одного аргумента a в соответствии с уравнением

$$\frac{df[V_n + at(V_n)]}{da} = 0. \quad (10.7)$$

Корень этого уравнения, максимизирующий функцию $f(V)$, обозначим a_n . Следующее приближение V_{n+1} вычисляется по формуле

$$V_{n+1} = V_n + a_n t(V_n).$$

Производится возврат к этапу 3.

В результате формируется последовательность приближений V_0, V_1, V_2, \dots . Вычислительный процесс заканчивается, когда будет достигнута точка V_n , в которой оценка градиента будет равна нулю (коэффициенты функции отклика становятся незначимыми).

Пример 1. Выполнить шаг крутого восхождения для функции отклика $y = -3x_1^2 - 2x_2^2$.

Решение.

Этап 1. Общий вид градиента функции $\nabla f(V)$:

$$\partial y / \partial x_1 = -6x_1, \quad \partial y / \partial x_2 = -4x_2; \quad \nabla f(V) = (-6x_1; -4x_2).$$

Длина вектора градиента:

$$|\nabla f(V)| = [(\partial y / \partial x_1)^2; (\partial y / \partial x_2)^2]^{0,5} = [36x_1^2; 16x_2^2]^{0,5}.$$

Единичный вектор t :

$$t = (t_1; t_2) = \nabla f(V) / |\nabla f(V)| = -(6x_1; 4x_2) / [36x_1^2 + 16x_2^2]^{0,5}.$$

Этап 2. Выбор начальной точки, например $V_0 = (5; 3)$.

Этап 3. Вычисление координат единичного вектора:

$$t(V_0) = -(30; 12) / [36 \cdot 25 + 16 \cdot 9]^{0,5} = -(30; 12) / [32,31] = (-0,93; -0,37).$$

Координаты точки V_1 при движении по направлению вектора \mathbf{t}

$$V_1 = V_0 + a \cdot \mathbf{t}(V_0) = (5; 3) + a \cdot (-0,93; -0,37) = (5 - a \cdot 0,93; 3 - a \cdot 0,37).$$

Функция отклика y в точке V_1 пространства двух переменных

$$y = -3 \cdot (5 - a \cdot 0,93)^2 - 2 \cdot (3 - a \cdot 0,37)^2.$$

Этап 4. Выбор шага a изменения координат текущей точки в соответствии с уравнением (10.7):

$$32,34 - 5,737 \cdot a = 0.$$

Следовательно, шаг $a = 5,637$.

Координаты точки V_1 после выполнения первого шага крутого восхождения $V_1 = V_0 + a \cdot \mathbf{t}(V_0) = (-0,242; 0,914)$.

Аналогично выполняется следующий шаг крутого восхождения.

Рассмотренный алгоритм применяют только для *нелинейных* функций. Если функция отклика является линейной, то выбор оптимального значения параметра a невозможен. В этом случае шаг выбирается исходя из эвристических предположений исследователя о виде функции отклика.

10.4. Особенности применения градиентной оптимизации совместно с методами планирования экспериментов

Применение методов планирования экспериментов вносит в типовую процедуру градиентных методов поиска свою специфику.

1. В задачах экспериментального исследования функция $f(V)$ обычно изначально неизвестна, ее вид выбирается относительно произвольно, а параметры устанавливаются по результатам эксперимента. На начальных этапах исследования трудоемкость решения задачи оптимизации можно снизить, применяя неполные полиномы k -того порядка или линейные полиномы

$$y' = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \dots \quad (10.8a)$$

$$+ \beta_{k-1,k} x_{k-1} x_k + \dots + \beta_{12\dots k} x_1 x_2 \dots x_k + \varepsilon;$$

$$y' = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (10.8б)$$

Таким образом, вместо самого градиента применяется его оценка. Оба вида полинома являются линейными относительно конкретного фактора. Количество членов полинома типа (10.8а) составляет 2^k , а для типа (10.8б) равно $k+1$. Теоретически оценки коэффициентов в точке оптимума должны стать равными нулю, что и будет признаком завершения поиска решения. Однако применение этих моделей может стать нерациональным в области, близкой к оптимуму, из-за больших относительных погрешностей в оценке коэффициентов указанных моделей. Поэтому для исследования области оптимума следует переходить к использованию полиномов более высокой степени.

2. Применение градиентных методов предполагает, что движение по градиенту может осуществляться в любом направлении изменения аргументов функции $f(V)$, т.е. ограничений на область допустимых значений аргументов нет. В практических задачах всегда существуют ограничения на значения параметров, поэтому при выборе направления движения следует учитывать это обстоятельство.

3. Значение градиента зависит от принятой системы перехода к кодированным значениям переменных, т.е. не является инвариантным к выбору центральной точки и интервала варьирования в формуле (10.4). Но знаки частных производных при переходе от одной системы координат к другой сохраняются. Поэтому направление перемещения в методе градиентного поиска не меняется при смене системы координат. Следовательно, в любой системе координат градиентный метод приведет к оптимуму, хотя скорость поиска и будет зависеть от выбранных значений центра и интервала варьирования переменных.

4. Рассмотренный выше способ определения шага крутого восхождения применяют только при описании поверхности отклика полными полиномами второй или более высокой степени. При анализе *линейных* функций определение шага изменения аргументов производится на основе неформальных процедур. Для полиномов (10.8а, б) шаг Δv_i^* изменения i -того

фактора относительно центра (в центре области планирования все нормализованные переменные равны нулю) определяется пропорционально соответствующей составляющей оценки градиента и величине интервала варьирования Δv_i

$$\Delta v_i^* = \Delta v_i \beta_i / [\beta_1^2 + \beta_2^2 + \dots + \beta_k^2]^{0,5}. \quad (10.9)$$

Новое значение основного уровня фактора $v_{i,1}$ в исходной шкале измерений составит величину $v_{i,1} = v_{i,0} + \Delta v_i^*$.

5. Применение метода крутого восхождения в его классическом виде предполагает вычисление градиента на каждом этапе. А это означает необходимость проведения достаточно большого количества опытов. Бокс и Уилсон предложили в 1951 г. модификацию метода крутого восхождения. Они рекомендуют на начальном этапе поиска применять линейные полиномы для описания функции отклика. Значение градиента оценивается в начальной точке, после чего пошаговое движение по градиенту продолжается до попадания в частный оптимум (до тех пор, пока значение функции отклика возрастает при переходе от точки к точке). В точке частного оптимума с помощью факторного эксперимента снова определяется градиент. И пошаговое движение начинается по новому направлению. Так продолжается до попадания в область глобального экстремума. Эта область не может быть адекватно описана линейным уравнением. Поэтому переходят к более точному описанию поверхности отклика на основе полиномов второго порядка и уточнению положения точки глобального оптимума. Построение плана для формирования полинома второй степени производится путем добавления некоторых точек к «ядру», уже сформированному для линейного приближения (такие планы получили наименование *композиционных*). В целом метод Бокса – Уилсона во многих случаях требует меньшего количества опытов возможно при несколько большем числе шагов.

6. Градиентные методы не обеспечивают гарантированного нахождения глобального оптимума при нарушении условия унимодальности функции отклика. Выбор начальной точки для крутого восхождения предопределяет

область поиска локального экстремума. Поэтому при наличии априорных сведений о возможности существования нескольких локальных экстремумов, целесообразно осуществить решение задачи оптимизации для нескольких вариантов задания исходных значений параметров.

7. Если эксперимент проводится на реальном объекте и требует больших затрат ресурсов, то поиск значений параметров может завершиться при получении удовлетворительных, а не оптимальных, значений функции отклика. Градиентный метод позволяет находить приемлемые решения и в этом случае.

Однако градиентный метод не всегда эффективен. Например, если поверхность функции отклика имеет овражный характер, то движение будет происходить с одного склона на другой с медленным продвижением к точке минимума. Для таких функций разработано несколько эвристических методов ускоренного продвижения вдоль оврага или гребня.

Контрольные вопросы

1. В чем состоит принцип метода движения по градиенту?
2. Назовите основные преимущества метода движения по градиенту?
3. Перечислите этапы движения по градиенту.
4. При каких условиях достигается область экстремума функции отклика?

11. КРИТЕРИИ ОПТИМАЛЬНОСТИ И ТИПЫ ПЛАНОВ

В настоящее время используется свыше 20 различных критериев оптимальности планов, которые подразделяются на две основные группы. К первой группе относят критерии, связанные с ошибками оценок коэффициентов, а ко второй – с ошибкой оценки поверхности отклика. Далее будут охарактеризованы только те критерии, которые наиболее часто применяются при решении задач оптимизации, описания поверхности отклика и оценки влияния факторов.

Критерии первой группы представляют интерес для задач оптимизации, выделения доминирующих (наиболее значимых) параметров на начальных этапах решения оптимизационных задач или для выявления несущественных параметров в задачах восстановления закономерности функционирования объекта. Геометрическое истолкование свойств ошибок коэффициентов связано со свойствами эллипсоида их рассеяния, определяемого математическим ожиданием и дисперсией значений ошибок. Пространственное расположение, форма и размер эллипсоида полностью зависят от плана эксперимента.

Критерию *D*-оптимальности соответствует минимальный объем эллипсоида рассеяния ошибок (минимум произведения всех дисперсий коэффициентов полинома). В соответствующем плане эффекты факторов максимально независимы друг от друга. Этот план минимизирует ожидаемую ошибку предсказания функции отклика. Критерию *A*-оптимальности соответствует план с минимальной суммарной дисперсией всех коэффициентов, а критерию *E*-оптимальности – план, в котором максимальная дисперсия коэффициентов будет минимальна.

Выбор критерия зависит от задачи исследования. Так, при изучении влияния отдельных факторов на поведение объекта применяют критерий *E*-оптимальности, а при поиске оптимума функции отклика – критерий *D*-оптимальности. Если построение *D*-оптимального плана вызывает затруднения, то можно перейти к *A*-оптимальному плану, построение которого осуществляется проще.

Критерии второй группы используются при решении задач описания поверхности отклика, определения ограничений на значения параметров. Основным здесь является критерий G -оптимальности, который позволяет построить план с минимальным значением наибольшей ошибки в описании функции отклика. Применение G -оптимального плана дает уверенность в том, что в области планирования нет точек с чрезмерно большой ошибкой описания функции.

Среди всех классов планов основное внимание в практической работе уделяется ортогональным и ротатабельным планам.

Ортогональным называется план, для которого выполняется условие парной ортогональности столбцов матрицы планирования, в частности, для независимых переменных

$$\sum_{i=1}^N x_{iu}x_{ju} = 0, \quad i \neq j, \quad i, j = 1, k,$$

где N – количество точек плана эксперимента, k – количество независимых факторов. При ортогональном планировании коэффициенты полинома определяются независимо друг от друга – вычеркивание или добавление слагаемых в функции отклика не изменяет значения остальных коэффициентов полинома. Для ортогональных планов эллипсоид рассеяния ориентирован в пространстве так, что направления его осей совпадают с направлениями координат пространства параметров.

Использование *ротатабельных* планов обеспечивает для любого направления от центра эксперимента равнозначность точности оценки функции отклика (постоянство дисперсии предсказания) на равных расстояниях от центра эксперимента. Это особенно важно при решении задач поиска оптимальных значений параметров на основе градиентного метода, так как исследователь до начала экспериментов не знает направление градиента и поэтому стремится принять план, точность которого одинакова во всех направлениях. В ряде случаев при исследовании поверхности отклика требуется униморфность модели, а именно, соблюдение постоянства значений дисперсии ошибки в некоторой области вокруг центра эксперимента.

Выполнение такого требования целесообразно в тех случаях, когда исследователь не знает точного расположения области поверхности отклика с оптимальными значениями параметров. Указанная область будет определена на основе упрощенной модели, полученной по результатам экспериментов.

По соотношению между количеством оцениваемых неизвестных параметров модели и количеством точек плана эксперимента все планы подразделяются на три класса: *ненасыщенные* – количество параметров меньше числа точек плана; *насыщенные* – обе величины одинаковы; *сверхнасыщенные* – количество параметров больше числа точек плана. Метод наименьших квадратов применяют только при ненасыщенном и насыщенном планировании, и он неприменим для сверхнасыщенного планирования.

Для некоторых планов важную роль играет свойство *композиционности*. Так, композиционные планы для построения полиномов второго порядка получают добавлением некоторых точек к планам формирования линейных функций. Это дает возможность в задачах исследования сначала попытаться построить линейную модель, а затем при необходимости, добавив наблюдения, перейти к моделям второго порядка, используя ранее полученные результаты и сохраняя при этом некоторое заданное свойство плана, например его ортогональность.

Между критериями оптимальности и методами построения оптимальных планов экспериментов существует жесткая связь. Построение планов производится или с использованием каталогов планов или с использованием непосредственно методов планирования экспериментов, что является непростой задачей и требует достаточно высокой квалификации исследователя в области ТПЭ.

Кроме рассмотренных критериев в планировании экспериментов вполне естественно применяется критерий минимума числа экспериментов, т.е. среди всех планов желательно выбирать такой, который требует минимального числа опытов при соблюдении требований к качеству оценки функции или ее параметров.

Как было отмечено выше, одной из областей применения ТПЭ является решение задач оптимизации, причем непосредственно для поиска оптимальных решений используются градиентные методы. Вычисление оценки градиента осуществляется на основе обработки экспериментальных данных. Хотя градиентный метод оптимизации не является составной частью ТПЭ, в целях удобства освоения материала далее приведено его краткое изложение.

Контрольные вопросы

1. В чем заключается смысл D -оптимальности плана эксперимента?
2. В чем заключается такое свойство плана эксперимента, как ортогональность?
3. Что обеспечивает такое свойство плана, как ротатабельность?
4. Что называют интерпретацией модели?
5. В какой момент обработки данных эксперимента проводят интерпретацию модели?
6. Что означает термин «значимость коэффициента модели»?
7. В чем заключается важность априорной информации при планировании эксперимента?

12. ПЛАНЫ ВТОРОГО ПОРЯДКА

Они позволяют сформировать функцию отклика в виде полного квадратичного полинома, который содержит большее число членов, чем неполный квадратичный полином, сформированный по планам первого порядка, и поэтому требуют большего числа выполняемых опытов. Полный квадратичный полином при $n=2$ содержит 6 членов

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2,$$

при $n=3$ – 11 членов

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2.$$

Известно, что для получения квадратичной зависимости каждый фактор должен фиксироваться как минимум на трех уровнях.

Для планов второго порядка область планирования может:

- быть естественной, то есть включать область планирования планов первого порядка и дополнительные точки (такие планы называются композиционными). Дополнительные точки могут выходить за область плана первого порядка – единичного гиперкуба. В этом случае опыты в них реализуются при установлении факторов за пределами варьирования. Это надо учитывать при определении области совместимости факторов;

- не выходить за пределы единичного гиперкуба, то есть для всех точек плана выполняется условие $|x_{iV}| \leq 1$;

- не выходить за пределы единичного гипершара, определяемого соотношением $x_1^2 + x_2^2 + \dots + x_n^2 \leq 1$.

Во втором и третьем случаях используют специальные приемы выполнения приведенных соотношений в плане. План с одной областью планирования можно перестроить в план другой областью планирования.

Если уже был ранее сформирован план ПФЭ, но точность его функции отклика не удовлетворяет, то мы можем достроить этот план до плана второго порядка (композиционный план) и сформировать функцию отклика в виде

полного квадратичного полинома без потери информации о ранее сделанных опытах.

12.1. Ортогональный центрально-композиционный план второго порядка

Ортогональным планом называется такой план, у которого матрица планирования X строится так, что бы матрица $C=X_tX$ оказалась диагональной. Используем этот подход и при построении планов второго порядка. План называется центральным, если все точки расположены симметрично относительно центра плана. ОЦКП – центральный симметричный ортогональный композиционный план [9, 14, 16].

В ОЦКП входят: ядро – план ПФЭ с $N_0=2^n$ точками плана, n_0 (одна для этого плана) центральная точка плана ($x_i=0, i=1,2,3,\dots,n$) и по две «звездные» точки для каждого фактора

$$x_i = \pm\alpha, \quad x_j = 0, \quad i=1,\dots,n, \quad j=1,\dots,n, \quad i \neq j,$$

α – плечо «звездных» точек.

При этом в каждой плоскости, содержащей ось Y и координатную ось i -того фактора (проходящей через центр плана), оказываются три значения фактора x_i ($-\alpha, 0, +\alpha$) и три соответствующих значения Y .

Общее количество точек в плане ОЦКП составляет

$$N = 2^n + 2n + n_0,$$

где для ОЦКП $n_0=1$.

При $n > 2$ в ОЦКП оказывается меньшее количество точек, чем в плане ПФЭ 3^n .

Число точек в плане:

n	2	3	4	5	6
ОЦКП	9	15	25	43	77
ПФЭ 3^2	9	27	81	243	729

Графическое представление ОЦКП для $n=3$ приведено на рис. 12.1.

Для ортогонального плана необходимо, чтобы выполнялось соотношение

$$\sum_{\nu=1}^N x_{i\nu} x_{j\nu} = 0.$$

Так как $x_{0\nu} = 1$, то для столбцов $j=1, 2, \dots, m+1$ должно выполняться условие

$$\sum_{\nu=1}^N x_{j\nu} = 0.$$

Это означает необходимость выполнения требования, чтобы сумма элементов любого столбца (кроме $j = 0$), включая столбцы, соответствующие квадратам фактора, была равна нулю. Это возможно, если члены столбцов, соответствующих квадратам факторов, преобразованы, иначе сумма квадратов факторов не может быть равна нулю.

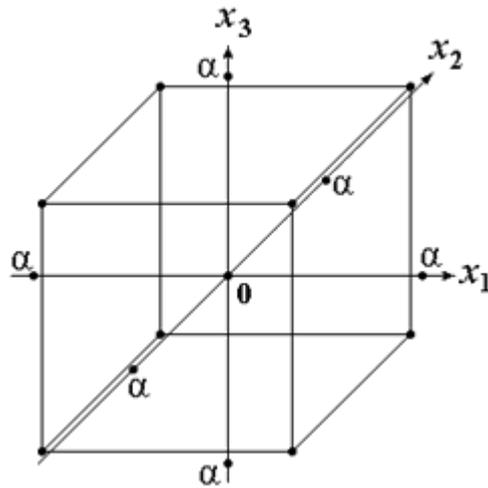


Рис. 12.1. ОЦКП при $n = 3$

Преобразование элементов этих столбцов осуществляется в виде

$$x'_{j\nu} = x_{j\nu}^2 - a,$$

где a – величина, зависящая от числа факторов.

Сумма элементов столбца, соответствующего квадратам факторов:

$$\sum_{\nu=1}^N x'_{j\nu} = \sum_{\nu=1}^N (x_{j\nu}^2 - a) = \sum_{\nu=1}^N x_{j\nu}^2 - N \cdot a = 0,$$

откуда

$$a = \frac{\sum_{\nu=1}^N x_{j\nu}^2}{N}.$$

В общем случае ортогональный центрально-композиционный план при трех факторах имеет следующий вид:

	U	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	$x_4' = x_1^2 - a$	$x_5' = x_2^2 - a$	$x_6' = x_3^2 - a$	Y
Точки плана ПФЭ 2^3 ($N_0 = 2^n$ точек)	1	+1	-1	-1	-1	+1	+1	+1	-1	$1-a$	$1-a$	$1-a$	Y_1
	2	+1	+1	-1	-1	-1	-1	+1	+1	$1-a$	$1-a$	$1-a$	Y_2
	3	+1	-1	+1	-1	-1	+1	-1	+1	$1-a$	$1-a$	$1-a$	Y_3
	4	+1	+1	+1	-1	+1	-1	-1	-1	$1-a$	$1-a$	$1-a$	Y_4
	5	+1	-1	-1	+1	+1	-1	-1	+1	$1-a$	$1-a$	$1-a$	Y_5
	6	+1	+1	-1	+1	-1	+1	-1	-1	$1-a$	$1-a$	$1-a$	Y_6
	7	+1	-1	+1	+1	-1	-1	+1	-1	$1-a$	$1-a$	$1-a$	Y_7
	8	+1	+1	+1	+1	+1	+1	+1	+1	$1-a$	$1-a$	$1-a$	Y_8
Звездные точки ($2n$ точек)	9	+1	$-\alpha$	0	0	0	0	0	0	$\alpha^2 - a$	$-a$	$-a$	Y_9
	10	+1	$+\alpha$	0	0	0	0	0	0	$\alpha^2 - a$	$-a$	$-a$	Y_{10}
	11	+1	0	$-\alpha$	0	0	0	0	0	$-a$	$\alpha^2 - a$	$-a$	Y_{11}
	12	+1	0	$+\alpha$	0	0	0	0	0	$-a$	$\alpha^2 - a$	$-a$	Y_{12}
	13	+1	0	0	$-\alpha$	0	0	0	0	$-a$	$-a$	$\alpha^2 - a$	Y_{13}
	14	+1	0	0	$+\alpha$	0	0	0	0	$-a$	$-a$	$\alpha^2 - a$	Y_{14}
Нулевая точка	15	+1	0	0	0	0	0	0	0	$-a$	$-a$	$-a$	Y_{15}
$\sum_{U=1}^N x_{iU}$	-	N	0	0	0	0	0	0	0	0	0	0	
$\sum_{U=1}^N x_{iU}^2$	-	N	$2^n + 2a^2$			2^n			$2^n(1-a)^2 + 2(\alpha^2 - a)^2 + a^2(2n-2) + n_0a^2$				

В ОЦКП каждый фактор фиксируется, в общем случае, на пяти уровнях ($-\alpha, -1, 0, 1, +\alpha$).

Для определения неизвестных « a » и « α » нужно сформировать и решить систему из двух уравнений. Одно из них для « a » мы записали ранее. Другое уравнение получим из условия ортогональности для столбцов x_4' и x_5' :

$$\sum_{U=1}^N x_{4U}' \cdot x_{5U}' = N_0(1-a)^2 - 4a(\alpha^2 - a) + a^2(2n-4) + n_0a^2 = 0.$$

После простейших преобразований с учетом того, что $N = N_0 + 2 \cdot n + n_0$ – общее число опытов в плане, получаем соотношение

$$\frac{N_0}{N} - 2 \frac{N_0 + 2a^2}{N} \cdot a + a^2 = 0.$$

Соотношение для a при $j=1, 2$ или 3 может быть записано как (см. план)

$$a = \frac{\sum_{j=1}^N x_{jV}^2}{N} = \frac{N_0 + 2 \cdot \alpha^2}{N}.$$

Подставив его в последнее уравнение, получаем

$$\frac{N_0}{N} - 2 \cdot a^2 + a^2 = 0,$$

откуда константа преобразования a

$$a = \sqrt{\frac{N_0}{N}} = \sqrt{\frac{2^n}{2^n + 2 \cdot n + n_0}},$$

тогда

$$\frac{N_0 + 2 \cdot \alpha^2}{N} = a = \sqrt{\frac{N_0}{N}}$$

и плечо звездных точек

$$\alpha = \sqrt{\frac{1}{2} (\sqrt{N \cdot N_0} - N_0)}.$$

Например, для ОЦКП при числе факторов $n=3$ имеем следующие параметры плана:

$$N_0 = 2^3 = 8, \quad N = 8 + 1 \cdot 3 + 1 = 15;$$

$$a = \sqrt{\frac{8}{15}} \approx 0,73; \quad \alpha = \sqrt{\frac{1}{2} (\sqrt{15 \cdot 8} - 8)} \approx 1,215;$$

$$1 - a = 0,27, \quad -a = -0,73, \quad \alpha^2 - a = 1,215^2 - 0,73 = 0,75.$$

Сам план принимает вид:

U	x_0	x_1	x_2	x_3	x_1x_2 x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	x'_4	x'_5	x'_6	Y
1	+1	-1	-1	-1	+1	+1	+1	-1	0,27	0,27	0,27	Y_1
2	+1	+1	-1	-1	-1	-1	+1	+1	0,27	0,27	0,27	Y_2
3	+1	-1	+1	-1	-1	+1	-1	+1	0,27	0,27	0,27	Y_3
4	+1	+1	+1	-1	+1	-1	-1	-1	0,27	0,27	0,27	Y_4
5	+1	-1	-1	+1	+1	-1	-1	+1	0,27	0,27	0,27	Y_5
6	+1	+1	-1	+1	-1	+1	-1	-1	0,27	0,27	0,27	Y_6
7	+1	-1	+1	+1	-1	-1	+1	-1	0,27	0,27	0,27	Y_7
8	+1	+1	+1	+1	+1	+1	+1	+1	0,27	0,27	0,27	Y_8
9	+1	-1,215	0	0	0	0	0	0	0,75	-0,73	-0,73	Y_9
10	+1	+1,215	0	0	0	0	0	0	0,75	-0,73	-0,73	Y_{10}
11	+1	0	-1,215	0	0	0	0	0	-0,73	0,75	-0,73	Y_{11}
12	+1	0	+1,215	0	0	0	0	0	-0,73	0,75	-0,73	Y_{12}
13	+1	0	0	-1,215	0	0	0	0	-0,73	-0,73	0,75	Y_{13}
14	+1	0	0	+1,215	0	0	0	0	-0,73	-0,73	0,75	Y_{14}
15	+1	0	0	0	0	0	0	0	-0,73	-0,73	-0,73	Y_{15}
$\sum_{U=1}^N x_{iU}$	N	0	0	0	0	0	0	0	0	0	0	
$\sum_{U=1}^N x_{iU}^2$	15	10,952			8			4,3727				

Очевидно, что план является ортогональным. В отличие от планов ПФЭ для ОЦКП сумма квадратов факторов разных столбцов не является одинаковой.

По результатам опытов плана формируется полином

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + b_4(x_1^2 - a) + b_5(x_2^2 - a) + b_6(x_3^2 - a)$$

Коэффициенты полинома $b_0, b_1, b_2, b_3, b_{12}, b_{13}, b_{23}, b_{123}, b_4, b_5, b_6$ определяются как

$$b_i = \frac{\sum_{U=1}^N x_{iU} Y_U}{\sum_{U=1}^N x_{iU}^2}$$

Можно преобразовать полином к виду

$$\hat{Y} = b'_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + b_4x_1^2 + b_5x_2^2 + b_6x_3^2,$$

где

$$b'_0 = b_0 - b_4 \cdot a + b_5 \cdot a - b_6 \cdot a.$$

Значения параметров ОЦКП при числе факторов n

n	2	3	4	5	6	7	8
α	1	1,215	1,414	1,596	1,761	1,909	2,045
a	0,667	0,73	0,8	0,86	0,91	0,946	0,968
N	9	15	25	43	77	143	273

При $n = 2$ ОЦКП совпадает с планом ПФЭ 2^3 . Звездные точки ОЦКП в этом случае лежат на границах варьирования факторов. Если точки плана ПФЭ 2^n всегда лежат на окружности (поверхности шара, гипершара), то точки плана ОЦКП не лежат на какой-либо одной окружности (поверхности шара, гипершара). План ОЦКП не является насыщенным. Так, например, для $n = 3$ полином имеет одиннадцать членов со своими коэффициентами, но для их определения используются пятнадцать опытов.

12.2. Пример плана ОЦКП для $n = 2$

Параметры плана $N_0=4$, $N=9$, $\alpha = 1$, $a = 2/3$, $1-a=1/3$, $-a=-2/3$, $\alpha^2 - a = -2/3$.

Использован рассмотренный ранее план ПФЭ 2^2 с добавленными опытами 5-9.

U	x_0	x_1	x_2	x_1x_2	$x'_3 = x_1^2 - a$	$x'_4 = x_2^2 - a$	Y	\hat{Y}	$ \hat{Y} - Y $
1	+1	-1	-1	+1	1/3	1/3	6	5,83	0,17
2	+1	+1	-1	-1	1/3	1/3	3	2,83	0,17
3	+1	-1	+1	-1	1/3	1/3	4	4,17	0,17
4	+1	+1	+1	+1	1/3	1/3	7	7,17	0,17
5	+1	-1	0	0	1/3	-2/3	5	5	0
6	+1	+1	0	0	1/3	-2/3	5	5	0
7	+1	0	-1	0	-2/3	1/3	1	1,33	0,33
8	+1	0	+1	0	-2/3	1/3	3	2,67	0,33
9	+1	0	0	0	-2/3	-2/3	2	2	0
$\sum_{U=1}^N x_{iU}^2$	9	6	6	4	2	2			

Коэффициенты полинома составляют

$$b_0 = \frac{\sum_{U=1}^9 x_0 Y_0}{\sum_{U=1}^9 x_0^2} = \frac{6+3+4+7+5+5+1+3+2}{9} = 4;$$

$$b_1 = \frac{-6+3-4+7-5+5+0 \cdot 1+0 \cdot 3+0 \cdot 2}{6} = 0;$$

$$b_2 = \frac{-6-3+4+7-0 \cdot 5+0 \cdot 5-1+3+0 \cdot 2}{6} = 0,67;$$

$$b_{12} = \frac{6-3-4+7}{4} = 1,5;$$

$$b_3 = \frac{1/3(6+3+4+7+5+5) - 2/3(1+3+2)}{6 \cdot (1/3)^2 + 3(2/3)^2} = 3;$$

$$b_4 = \frac{1/3(6+3+4+7+1+3) - 2/3(5+5+2)}{2} = 0.$$

Полином принимает вид

$$\hat{Y} = 4 + 0 \cdot x_1 + 0,67x_2 + 3(x_1^2 - 0,67) + 0 \cdot (x_2^2 - 0,67) + 1,5x_1x_2 = 2 + 0,67x_2 + 3x_1^2 + 1,5x_1x_2.$$

(Ранее по плану ПФЭ 2^2 был сформирован полином $\hat{Y} = 5 + 0,5x_2 + 1,5x_1x_2$.)

Рассчитанные значения \hat{Y} по полиному приведены в плане. Также приведены величины $|\hat{Y}_V - Y_V|$, подтверждающие достаточно высокую точность полинома. Так в центральной точке плана, в отличие от случая применения плана ПФЭ 2^2 , расхождений нет.

12.3. Ротатабельные планы

Ротатабельные планы – это планы, у которых точки плана располагаются на окружностях (сферах, гиперсферах). У ротатабельного плана первого порядка точки плана располагаются на одной окружности (сфере, гиперсфере) с радиусом R :

$$\sqrt{\sum_{i=1}^n x_{iV}^2} = \text{const} = R,$$

где $V=1, \dots, N$ – номер точки плана, $i=1, \dots, n$ – номер фактора.

В таком случае точность оценивания функции отклика по любому направлению факторного пространства (для всех точек плана) одинаковая.

Ротатабельный план может быть симметричным, когда точки плана располагаются симметрично относительно друг друга. Рассмотренный ранее план ПФЭ 2^n – ротатабельный симметричный план первого порядка.

У ротатабельных планов второго порядка точки плана располагаются на двух концентрических гиперсферах с радиусами R_1 и R_2 . В таких планах

$$\sqrt{\sum_{i=1}^n x_{iV}^2} = \text{const}_1 = R_1$$

для $V=1, \dots, N_0$ и

$$\sqrt{\sum_{i=1}^n x_{iW}^2} = \text{const}_2 = R_2$$

для $W=1, \dots, n_0$, где V и W – текущие номера точек плана в двух подмножествах опытов N_0 и n_0 из их общего количества N , относящихся к двум разным концентрическим сферам. Одна из сфер может быть вырожденной, когда $R_2=0$. Рассмотренный ранее ортогональный центрально-композиционный план второго порядка (ОЦКП) не является ротатабельным планом, так как его точки лежат на трех концентрических окружностях (сферах, гиперсферах). При $n=2$ это очевидно из рис. 12.2. «Звездные» точки плана и точки плана ПФЭ 2^n лежат на разных окружностях.

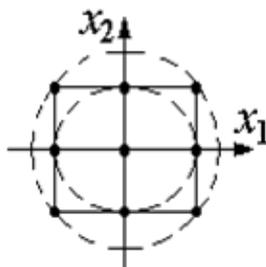


Рис. 12.2. Расположение точек ОЦКП на трех окружностях

Ротатабельный план может быть ортогональным, если выполняется условие

$$\sum_{U=1}^N x_{iU} \cdot x_{jU} = 0,$$

где $i=1, \dots, m$, $j=1, \dots, m$, $m > n$, $i \neq j$ – номера столбцов плана.

12.4. Ротатабельный ортогональный центрально-композиционный план

Ротатабельный ортогональный центрально-композиционный план (РОЦКП) строится аналогично рассмотренному ранее ОЦКП. К

использованному в качестве ядра плану ПФЭ 2^n добавляются «звездные» точки – по две на каждый фактор и несколько точек в центре плана. «Звездные» точки должны располагаться на поверхности гиперсферы с радиусом R , на которой лежат и точки плана ПФЭ 2^n , то есть величина плеча «звездных» точек α должна равняться радиусу R . Это может быть обеспечено, при выполнении условия ортогональности, только при соответствующем выборе числа наблюдений в центральной (нулевой) точке плана n_0 . Для РОЦКП n_0 зависит от числа факторов n . Напомним, что в ОЦКП $n_0 = 1$ для любого числа n .

Радиус сферы, на которой лежат точки плана ПФЭ 2^n при двух уровнях варьирования факторов с диапазоном ± 1 составляет (рис. 12.3).

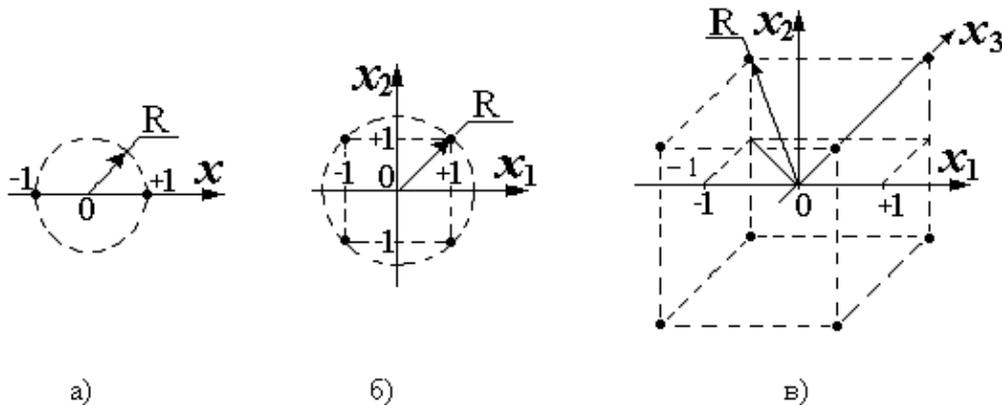


Рис. 12.3. Радиус окружности (сферы), на которой лежат точки плана ПФЭ 2^n при диапазоне варьирования факторов от -1 до $+1$:

$$\text{а) } - n=1, R = \sqrt{1} = 1;$$

$$\text{б) } - n=2, R = \sqrt{2} = 1,414;$$

$$\text{в) } - n=3, R = \sqrt{3} = 1,732;$$

$$R = \sqrt{n}$$

Таким образом, при построении РОЦКП с ядром из плана ПФЭ 2^n плечо «звездных» точек определяется числом факторов

$$\alpha = \sqrt{n}.$$

Ранее при определении параметров ортогонального композиционного плана второго порядка с ядром из плана ПФЭ 2^n было получено

$$\alpha = \sqrt{\frac{1}{2}(\sqrt{N \cdot N_0} - N_0)},$$

где $N_0 = 2^n$ – число точек плана ПФЭ,

$N = N_0 + 2n + n_0$ – полное число точек композиционного плана второго порядка,

$a = \sqrt{N_0/N}$ – константа преобразования элементов столбцов, соответствующих квадратам факторов.

В этом случае для РОЦКП число наблюдений в центре плана

$$n_0 = \frac{4n^2}{2^n} + 2n.$$

Если n_0 не целое, то при практическом построении плана его округляют до целого, но свойство ортогональности плана нарушается.

Параметры РОЦКП в зависимости от числа факторов:

n	1	2	3	4	5	6	7	8
$\alpha = \sqrt{n}$	1	1,414	1,732	2	2,236	2,45	2,646	2,83
$n_0 = \frac{4n^2}{2^n} + 2n$	4	8	10,5	12	13,13	14,25	15,53	17
$N_0 = 2^n$	2	4	8	16	32	64	128	256
N	8	16	24,5	36	55,13	90	157,55	289
$a = \sqrt{N_0/N}$	0,5	0,5	0,574	0,67	0,76	0,84	0,9	0,94

Для РОЦКП рекомендуется принимать

$$\alpha' = N_0^{\frac{1}{4}} = (2^n)^{\frac{1}{4}} = 2^{\frac{n}{4}},$$

тогда

$$n_0' = 4 - 2n + 2^{\frac{n+4}{2}}.$$

Параметры РОЦКП:

N	1	2	3	4	5	6	7	8
$\alpha' = 2^{\frac{n}{4}}$	1,189	1,414	1,682	2	2,378	2,838	3,364	4
$n_0' = 4 - 2n + 2^{\frac{n+4}{2}}$	7,66	8	9,31	12	16,63	24	33,25	52
N'	11,66	16	23,31	36	58,63	100	177,25	324
a'	0,414	0,5	0,586	0,57	0,739	0,8	0,85	0,889

12.5. Пример ротатабельного ортогонального центрально-композиционного плана для $n = 2$

Параметры плана:

$$\alpha = \sqrt{2} = 1,414, \quad n_0 = 8, \quad N_0 = 2^2 = 4, \quad N = 2^2 + 2 \cdot 2 + n_0 = 16,$$

$$a = \sqrt{\frac{N_0}{N}} = 0,5, \quad 1 - a = 0,5, \quad -a = -0,5, \quad \alpha^2 - a = 2 - 0,5 = 1,5.$$

Нет необходимости проводить восемь раз (точки с 9 по 16) опыты в центре плана. Достаточно провести этот опыт один раз и записать результат во все восемь строк. Строки сокращать нельзя, так как нарушается свойство ортогональности, и коэффициенты полинома будут определены неверно.

Коэффициенты квадратичного полинома рассчитаются, как и ранее.

Использован рассмотренный ранее план ПФЭ 2^2 с добавленными опытами 5-16.

	U	x_0	x_1	x_2	$x_1 x_2$	$x'_3 = x_1^2 - a$	$x'_4 = x_2^2 - a$	Y	\hat{Y}	$ \hat{Y} - Y $
ПФЭ 2^2	1	+1	-1	-1	+1	0,5	0,5	6	5,146	0,854
	2	+1	+1	-1	-1	0,5	0,5	3	2,146	0,854
	3	+1	-1	+1	-1	0,5	0,5	4	3,35	0,65
	4	+1	+1	+1	+1	0,5	0,5	7	6,35	0,65
Звездные точки	5	+1	-1,414	0	0	1,5	-0,5	5	5,75	0,75
	6	+1	+1,414	0	0	1,5	-0,5	5	5,75	0,75
	7	+1	0	-1,414	0	-0,5	1,5	1	1,9	0,9
	8	+1	0	+1,414	0	-0,5	1,5	3	3,6	0,6
Точки в центре плана	9	+1	0	0	0	-0,5	-0,5	2	2	0
	10	+1	0	0	0	-0,5	-0,5	2	2	0
	11	+1	0	0	0	-0,5	-0,5	2	2	0
	12	+1	0	0	0	-0,5	-0,5	2	2	0
	13	+1	0	0	0	-0,5	-0,5	2	2	0
	14	+1	0	0	0	-0,5	-0,5	2	2	0
	15	+1	0	0	0	-0,5	-0,5	2	2	0
	16	+1	0	0	0	-0,5	-0,5	2	2	0
$\sum_{i=1}^N x_{iU}$	16	0	0	0	0	0				
$\sum_{i=1}^N x_{iU}^2$	16	8	8	4	8	8				

$$\varepsilon_0 = \frac{6+3+4+7+5+5+1+3+2 \cdot 8}{16} = \frac{50}{16} = 3,125,$$

$$\varepsilon_1 = \frac{-6+3-4+7-5 \cdot 1,414+5 \cdot 1,414+10 \cdot 0}{4 \cdot 1^2 + 2(1,414)^2 + 10 \cdot 0^2} = \frac{0}{8} = 0, \quad \varepsilon_2 = 0,6035,$$

$$\varepsilon_{12} = \frac{6-3-4+7+12 \cdot 0}{4 \cdot 1^2} = 1,5,$$

$$\varepsilon_{11} = \frac{0,5(6+3+4+7)+1,5(5+5)-0,5(1+3+8 \cdot 2)}{14 \cdot 0,5^2 + 2 \cdot 1,5^2} = \frac{15}{8} = 1,875,$$

$$\varepsilon_{22} = \frac{0,5(6+3+4+7)-0,5(5+5)+1,5(1+3)-0,5 \cdot 2 \cdot 8}{8} = 0,375.$$

Полином принимает следующий вид:

$$\hat{Y}_{\text{РОЦКП}} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11}(x_1^2 - a) + b_{11}(x_2^2 - a) = 3,125 + 0 \cdot x_1 + 0,6035 x_2 + 1,5 x_1 x_2 + 1,875(x_1^2 - 0,5) + 0,375(x_2^2 - 0,5) = 2 + 0 \cdot x_1 + 0,6035 x_2 + 1,5 x_1 x_2 + 1,875 x_1^2 + 0,375 x_2^2$$

Рассчитанные значения функции и расхождения с опытными данными представлены в предпоследнем и последнем столбцах плана.

Ранее для ОЦКП, при несколько отличающейся поверхности функции, был получен близкий полином в виде

$$\hat{Y}_{\text{ОЦКП}} = 2 + 0 \cdot x_1 + 0,67 x_2 + 1,5 x_1 x_2 + 3 x_1^2 + 0 \cdot x_2^2.$$

Для $n = 2$ число членов квадратичного полинома составляет шесть. В ОЦКП и РОЦКП необходимо провести девять отличающихся опытов при пяти уровнях варьирования факторов. Поэтому ОЦКП и РОЦКП – ненасыщенные планы. Такое число экспериментальных точек может быть использовано для построения, например, кубических полиномов.

12.6. Планы второго порядка с единичной областью планирования

Так как ОЦКП и РОЦКП – композиционные планы, то при естественной области планирования «звездные» точки могут выходить за пределы единичного гиперкуба и единичного гипершара. Для вписывания плана в область единичного гипершара необходимо изменить значения факторов путем умножения их на коэффициент

$$C = \frac{1}{\sqrt{n}}.$$

Так, при $n = 2$,

$$C_{\text{РОЦКП}} = C_{\text{ОЦКП}} = \frac{1}{\sqrt{2}} = 0,707.$$

Значения факторов в ОЦКП и РОЦКП при переходе от естественной области планирования к единичному гипершару при $n = 2$:

U	ОЦКП		РОЦКП	
	x_1	x_2	x_1	x_2
1	-0,707	-0,707	-0,707	-0,707
2	+0,707	-0,707	+0,707	-0,707
3	-0,707	+0,707	-0,707	+0,707
4	+0,707	+0,707	+0,707	+0,707
5	-0,707	0	-1	0
6	+0,707	0	+1	0
7	0	-0,707	0	-1
8	0	+0,707	0	+1
9	0	0	0	0
10	-		0	0
11	-		0	0
12	-		0	0
13	-		0	0
14	-		0	0
15	-		0	0
16	-		0	0

Могут использоваться ротатабельные планы с точками плана в вершинах других, кроме квадрата (куба, суперкуба), правильных многогранников, вписанных в область единичного круга (шара, гипершара). В ротатабельном плане на основе N_0 -угольника присутствуют N_0 отличающихся точек на окружности с радиусом $R_1=1$ и n_0 совпадающих точек в центре плана с радиусом $R_2=0$. При $n=2$ для квадратичного полинома при шести его членах число отличающихся точек плана должно быть не менее шести. В планах на основе пятиугольника (шестиугольника или семиугольника) присутствуют 6 (7 или 8) отличающихся точек, что меньше, чем в ОЦКП и РОЦКП, у которых 9 отличающихся точек. При соответствующем выборе многоугольника можно

сформировать насыщенный ротатабельный план второго порядка. Значения факторов в точках плана определяются типом многоугольника.

12.7. Ротатабельный план на основе правильного многоугольника при $n=2$

U	x_0	x_1	x_2	$x_3 = x_1 x_2$	$x_4 = x_1^2 - a$	$x_5 = x_2^2 - a$
1	1	$\cos 0$	$\sin 0$	$\frac{\sin 0}{2}$	$\cos^2 0 - \frac{N_0}{2N}$	$\sin^2 0 - \frac{N_0}{2N}$
2	1	$\cos \frac{2\pi}{N_0}$	$\sin \frac{2\pi}{N_0}$	$\frac{\sin \frac{4\pi}{N_0}}{2}$	$\cos^2 \frac{2\pi}{N_0} - 0,5 \frac{N_0}{N}$	$\sin^2 \frac{2\pi}{N_0} - 0,5 \frac{N_0}{N}$
...
V	1	$\cos \frac{2\pi(V-1)}{N_0}$	$\sin \frac{2\pi(V-1)}{N_0}$	$\frac{\sin \frac{4\pi(V-1)}{N_0}}{2}$	$\cos^2 \frac{2\pi(V-1)}{N_0} - 0,5 \frac{N_0}{N}$	$\sin^2 \frac{2\pi(V-1)}{N_0} - 0,5 \frac{N_0}{N}$
N
N_0	1	$\cos(2\pi - \frac{2\pi}{N_0})$	$\sin(2\pi - \frac{2\pi}{N_0})$	$\frac{\sin(4\pi - \frac{4\pi}{N_0})}{2}$	$\cos^2(2\pi - \frac{2\pi}{N_0}) - 0,5 \frac{N_0}{N}$	$\sin^2(2\pi - \frac{2\pi}{N_0}) - 0,5 \frac{N_0}{N}$
1	1	0	0	0	$-0,5 \frac{N_0}{N}$	$-0,5 \frac{N_0}{N}$
...
W	1	0	0	0	$-0,5 \frac{N_0}{N}$	$-0,5 \frac{N_0}{N}$
...
x_0	1	0	0	0	$-0,5 \frac{N_0}{N}$	$-0,5 \frac{N_0}{N}$
$\sum_{U=1}^N x_{iU}$	N	0	0	0	0	0
$\sum_{U=1}^N x_{iU}^2$	N	$0,5N_0$	$0,125N_0$	$0,375N_0 - 0,25 \frac{N_0^2}{N} = 0,25N_0$		

Константа преобразования элементов столбцов, соответствующих квадратам факторов, для всех подобных планов составляет

$$a = \frac{\sum_{U=1}^N x_{jU}^2}{N} = 0,5 \frac{N_0}{N}.$$

Смотри, например, для столбцы $i=1$ или 2 приведенного плана.

Соотношение $\frac{N_0}{N}$ может быть определено из уравнения выполнения условия ортогональности столбцов x'_4 и x'_5 :

$$\sum_{\nu=1}^N x'_{4\nu} \cdot x'_{5\nu} = 0.$$

После несложных преобразований оно сводится к требованию

$$0,125N_0 - 0,25 \frac{N_0^2}{N} = 0,$$

что выполняется при условии в таких планах

$$\frac{N_0}{N} = \frac{1}{2},$$

и следовательно $N_0 = n_0 = 0,5N$.

Таким образом, число точек в центре плана для всех подобных планов равно числу точек на поверхности единичного гипершара и определяется типом использованного многогранника [9, 12, 14].

Константа преобразования для всех подобных планов составляет $a=0,25$.

Например, в ротатабельном плане при $n=2$ на основе правильного шестиугольника присутствуют 7 отличающихся точек: $N_0=6$ точек на единичной окружности и $n_0=6$ совпадающих точек в центре плана (рис. 12.4).

Здесь при построении плана первый фактор варьируется на пяти уровнях, а второй – на трех уровнях.

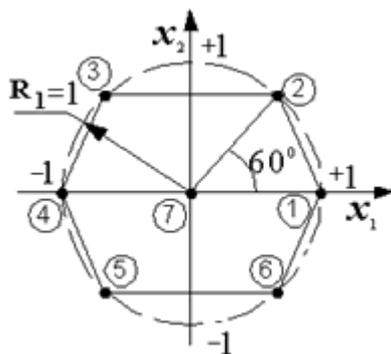


Рис. 12.4. Ротатабельный план при $n=2$ на основе правильного шестиугольника

Ротатабельный план при $n=2$ на основе шестиугольника:

	U	x_1	x_2	$\sum_{i=1}^n x_{iU}^2 = R$	$x_3 = x_1 x_2$	$x'_4 = x_1^2 - 0,25$	$x'_5 = x_2^2 - 0,25$
N_0	1	1	0	1,0	0	0,75	-0,25
	2	0,5	0,866	1,0	0,433	0	0,5
	3	-0,5	0,866	1,0	-0,433	0	0,5
	4	-1	0	1,0	0	0,75	-0,25
	5	-0,5	-0,866	1,0	0,433	0	0,5
	6	0,5	-0,866	1,0	-0,433	0	0,5
n_0	7	0	0	0,0	0	-0,25	-0,25
	8	0	0	0,0	0	-0,25	-0,25
	9	0	0	0,0	0	-0,25	-0,25
	10	0	0	0,0	0	-0,25	-0,25
	11	0	0	0,0	0	-0,25	-0,25
	12	0	0	0,0	0	-0,25	-0,25
$\sum_{U=1}^N x_{iU}$	0	0			0	0	0
$\sum_{U=1}^N x_{iU}^2$	3,0	3,0			0,75	1,5	1,5

Существуют ротатабельные планы, где оба радиуса не нулевые. При этом количество точек на каждой поверхности и отношение радиусов связаны.

Числа точек окружностей ротатабельного плана и отношение их радиусов:

Число точек внешней окружности (N_0)	6	7	8	7	8	8
Число точек внутренней окружности (n_0)	5	5	5	6	6	7
Отношение радиусов окружностей (R_2/R_1)	0,204	0,267	0,304	0,189	0,25	0,176

Пример такого плана при $n=2$, $N_0=8$, $n_0=6$, $R_2 / R_1=0,25$.

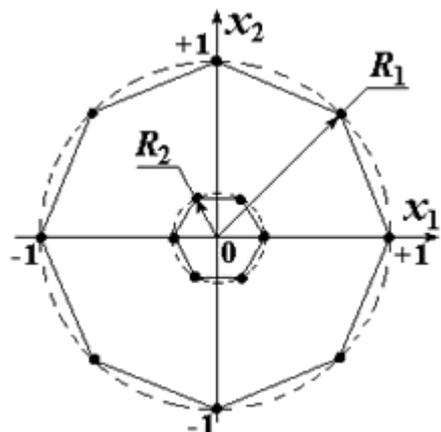


Рис. 12.5. Ротатабельный план с двумя невырожденными окружностями

Контрольные вопросы

1. С какой целью применяются планы второго порядка?
2. Какой может быть область планирования для планов второго порядка?
3. Как определяется число опытов для ортогонального центрально-композиционного плана второго порядка?
4. Как определяются константа преобразования и число звездных точек для ортогонального центрально-композиционного плана второго порядка?
5. Какие планы называются ротатабельными?
6. Как определяется число опытов для ротатабельного ортогонального центрально-композиционного плана второго порядка?

13. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

13.1. Парная корреляция

Методы корреляционного и регрессионного анализа имеют существенное значение при решении многих вопросов. При решении широкого круга задач прежде всего необходимо знать, связаны или не связаны между собой две или более случайные величины. Решение такого рода задач в инженерной практике, в частности, возникает при необходимости выявления зависимости между некоторой статической или динамической возмущающей силой и наблюдаемой реакцией исследуемой физической системы.

Основная задача корреляционного анализа – выявление значимости связи между значениями различных случайных величин. Величины, значения которых не зависят от того, какие значения получили некоторые другие величины, называются независимыми от них.

Зависимость между величинами (в том числе и случайными), при которых каждому значению одной величины (аргумента) отвечает одно или несколько вполне определенных значений другой, называют соответственно однозначной или многозначной функциональной зависимостью.

Зависимость между величинами, при которой каждому значению одной величины отвечает с соответствующей вероятностью множество возможных значений другой, называют вероятностной (стохастической, статистической). В общем случае вероятностной связи при изменении значения одной величины изменяется условный закон распределения другой.

Если при наличии вероятностной зависимости между двумя величинами с изменением значения одной величины изменяется только математическое ожидание второй (и наоборот), а дисперсия, области возможных значений и тип закона распределения остаются неизменными, то для таких величин характерна корреляционная зависимость.

Примерами корреляционной связи являются зависимости: между пределами прочности и текучести стали определенной марки, между погрешностью размера и погрешностью формы поверхности детали,

обработанной определенным методом, между температурой испытания и ударной вязкостью стали, между усилием прижима ролика и шероховатостью накатанной детали. В первых двух примерах имеет место корреляционная связь между двумя откликами, а в третьем и четвертом – между фактором, который является случайной величиной в связи с погрешностью измерения, и откликом.

Предположим, что X и Y – случайные переменные, имеющие нормальное распределение. Силу линейной статистической связи между ними можно оценить коэффициентом корреляции ρ , который принимает значения в интервале $(-1, +1)$ и не зависит от выбора канала отсчета и единиц величин X и Y :

$$\rho = M \left[\frac{X - M(X)}{\sqrt{D(X)}} \right] \left[\frac{Y - M(Y)}{\sqrt{D(Y)}} \right].$$

Чем больше отличается от нуля коэффициент корреляции, тем сильнее зависимость между величинами X и Y . Коэффициент корреляции независимых величин равен нулю. Однако обратное утверждение не всегда является верным, потому что на коэффициент корреляции оказывает влияние также отклонение от линейности связи.

Характеристикой нелинейной корреляционной связи является корреляционное отношение

$$\eta^2 = \frac{M[M(Y/X = x) - M(Y)]^2}{M[Y - M(Y)]^2},$$

где $M(Y/X = x)$ – условное математическое ожидание случайной переменной Y , рассматриваемое как функция x .

Оценкой коэффициента корреляции является значение коэффициента r . Для его вычисления необходимо знать оценки математических ожиданий $M(X)$ и $M(Y)$, а также дисперсий $D(X)$ и $D(Y)$. Если выполнено m наблюдений, коэффициент корреляции рассчитывается по следующей формуле:

$$r = \frac{\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y})}{(m-1) S(X) S(Y)}. \quad (13.1)$$

При относительно небольшом m удобно пользоваться следующими формулами:

$$\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^m X_i Y_i - \frac{1}{m} \sum_{i=1}^m X_i \sum_{i=1}^m Y_i; \quad (13.2)$$

$$(m-1) S^2(X) = \sum_{i=1}^m X_i^2 - \frac{1}{m} \left(\sum_{i=1}^m X_i \right)^2;$$

$$(m-1) S^2(Y) = \sum_{i=1}^m Y_i^2 - \frac{1}{m} \left(\sum_{i=1}^m Y_i \right)^2.$$

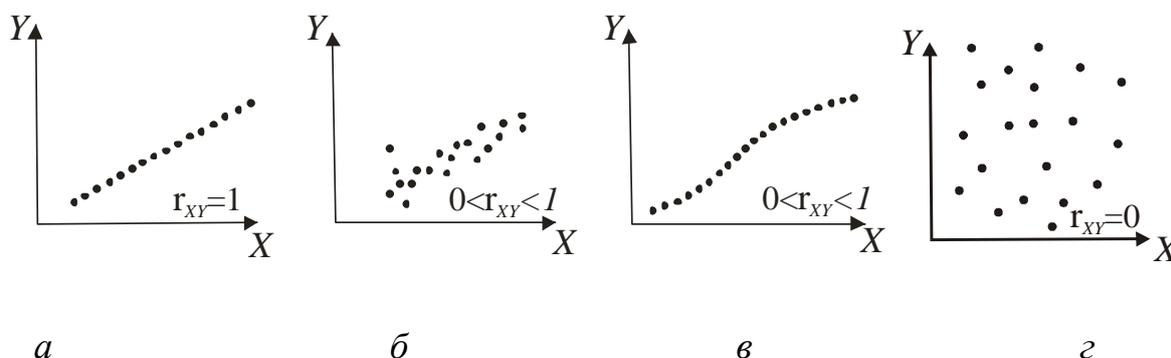


Рис. 13.1. Примеры различных видов корреляции:

a – идеальная линейная корреляция;

б – линейная корреляция с умеренным рассеянием; *в* – нелинейная корреляция;

г – отсутствие корреляции; r_{XY} – коэффициент корреляции

Пару случайных чисел X_i, Y_i можно изобразить графически в виде точки с координатами (X_i, Y_i) (рис. 13.1).

На практике при небольших значениях m для вычисления выборочного коэффициента корреляции r используют поле корреляции (рис. 13.2).

По осям координат откладываются интервалы изменения переменных и наносится координатная сетка. Каждую пару переменных из данной выборки изображают точкой в соответствующей клетке.

Такое изображение называют полем корреляции. На рис. 13.2 показано поле корреляции для 106 совместных измерений предела прочности σ_B и предела текучести σ_T стали 30ХГСА [5].

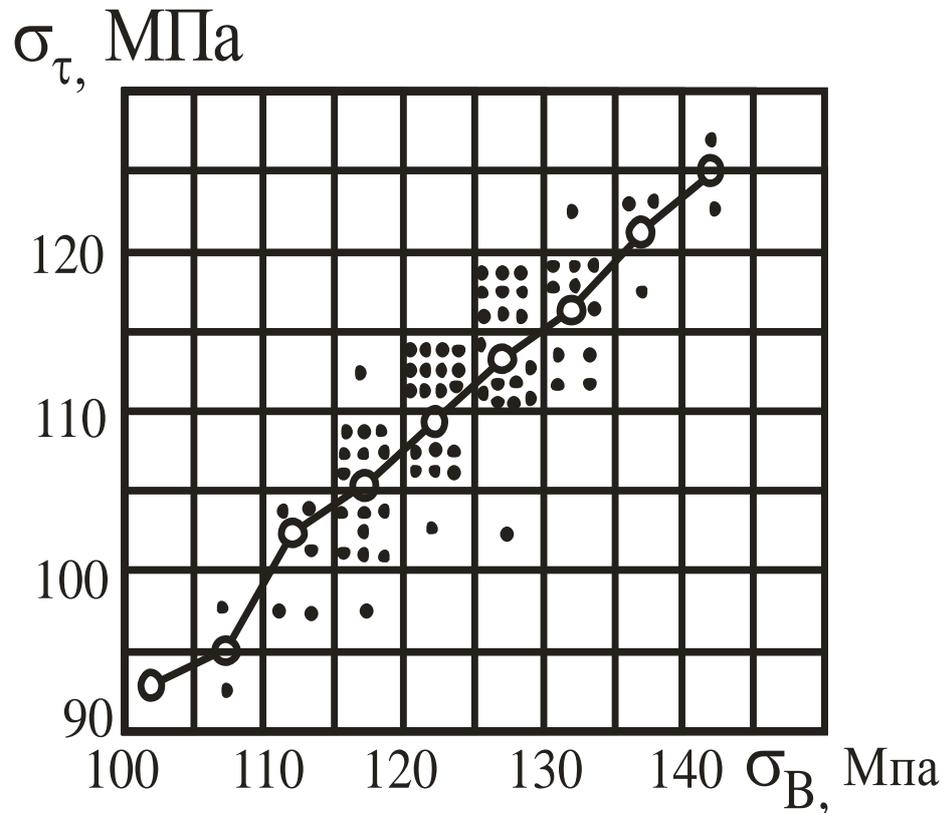


Рис. 13.2. Поле корреляции совместных измерений предела прочности σ_B и предела текучести σ_τ стали 30ХГСА

Поле корреляции позволяет построить корреляционную таблицу (табл. 13.1).

В ячейки, образованные пересечением строк и столбцов, заносятся частоты m_{XY} и попадания пар значений (X_i, Y_i) в соответствующие интервалы поля корреляции. В первом столбце и первой строке корреляционной таблицы указывают середины интервалов изменения случайных величин, а в последних – суммы частот m_{XY} по строкам и столбцам (m_X и m_Y соответственно).

Представленные формулы (13.2) можно преобразовать следующим образом:

$$\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^m x_i y_i m_{XY} - \frac{1}{m} \sum_{i=1}^{m_1} x_i m_X \sum_{i=1}^{m_2} x_i m_Y .$$

Корреляционная таблица для σ_B и σ_T

σ_T	92,5	97,5	102,5	107,5	112,5	117,5	122,5	127,5	Итого
σ_B	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
102,5 (2)	1								1
107,5 (3)	1	1							2
112,5 (4)		2	4	2					8
117,5 (5)		1	7	11	1				20
122,5 (6)			1	13	13				27
127,5 (7)			1	2	15	9			27
132,5 (8)					4	10	1		15
137,5 (9)						1	3		4
142,5 (10)							1	1	2
Итого	2	4	13	28	33	20	5	1	106

$$mS^2(X) = \sum_{i=1}^{m_1} x_i^2 m_X - \frac{1}{m} \left(\sum_{i=1}^{m_1} x_i m_X \right)^2 ;$$

$$mS^2(Y) = \sum_{i=1}^{m_2} \bar{y}_i^2 m_Y - \frac{1}{m} \left(\sum_{i=1}^{m_2} Y_i m_Y \right)^2 ,$$

где x_i , y_i – середины соответствующих интервалов изменения величин;

m_1 – число столбцов; m_2 – число строк корреляционной таблицы;

$m = \sum m_{XY} = \sum m_X = \sum m_Y$.

Если диапазоны изменения X и Y разделены на равные интервалы (Δx и Δy), то вместо натуральных значений можно использовать целочисленные коды. Для этого нужно выполнить линейное преобразование x и y по формулам

$$x_{ki} = \frac{x_i - x_{\min}}{\Delta x}; \quad y_{ki} = \frac{y_i - y_{\min}}{\Delta y} - \frac{x_{\min}}{\Delta x}.$$

Тогда коды x_i примут значения $x_{ki} = 0, 1, 2, \dots, i$, а коды y_i будут сдвинуты на целое число:

$$\Delta = y_{\min} / \Delta y - x_{\min} / \Delta x.$$

Проверка значимости коэффициента корреляции производится при помощи критерия Стьюдента. Наблюдаемое значение критерия

$$t_{\text{набл}} = r \sqrt{(m-2)/(1-r^2)}. \quad (13.3)$$

Критическое значение критерия определяется из таблицы критических значений Стьюдента по принятому значению доверительной вероятности P и числу степеней свободы $f = m - 2$. Поэтому при выборе $t_{\text{кр}}$ следует принимать $m = f + 1$. Если $|t_{\text{набл}}| < t_{\text{кр}}$, то принимается нулевая гипотеза $\rho = 0$. В противном случае она отклоняется.

Выборочное корреляционное отношение вычисляется по формуле

$$\eta^2 = S^2(y/x) / S^2(y), \quad (13.4)$$

где $S^2(y)$ вычисляется по вышеприведенным формулам; $S^2(y/x)$ – условная дисперсия:

$$S^2(y/x) = \frac{1}{m} \sum [\bar{y}(x) - \bar{y}]^2 m_x;$$

$$\bar{y}(x) = \frac{1}{m_x} \sum y m_{xy}.$$

Корреляционное отношение η^2 связано с ρ^2 следующим образом: $0 \leq \rho^2 \leq \eta^2 \leq 1$. В случае линейной зависимости между переменными будет соблюдаться равенство: $\rho^2 = \eta^2$. Разность $\eta^2 - \rho^2$ может служить показателем нелинейной связи.

Пример 1. На основании данных корреляционной таблицы (см. табл. 13.1) необходимо определить выборочный коэффициент корреляции r между σ_{τ} и $\sigma_{\text{в}}$.

Решение. Кодом для $\sigma_{\text{в}}$ выбираем x , а для σ_{τ} – y . Тогда $\Delta = 2$. Коды записаны в корреляционной таблице под действительными значениями в скобках.

$$\sum_{i=1}^{106} (X_i - \bar{X})(Y_i - \bar{Y}) = 2569 - \frac{1}{106} 667 \cdot 384 = 152,698;$$

$$mS^2(X) = 4429 - \frac{1}{106} 667^2 = 231,934 \quad S(x) = 1,479;$$

$$mS^2(Y) = 1566 - \frac{1}{106} 384^2 = 174,906 \quad S(y) = 1,285.$$

Из рассчитанных значений по формуле (13.1)

$$r = \frac{152,698}{(105 \cdot 1,479 \cdot 1,285)} = 0,763.$$

Пример 2. Для условий рассмотренного выше примера определить корреляционное отношение. Значения $\bar{y}(x)$ для соответствующих значений кодов $\sigma_{\text{в}}$ (первый столбец табл. 13.1) составляют:

x	2	3	4	5	6	7	8	9	10
$\bar{y}(x)$	0	0,5	2	2,6	3,44	4,19	4,8	5,75	6,5

Решение.

$$\bar{y} - \frac{1}{m} \sum y_i m_Y = 3,62; S^2(y/x) = 1,317; \eta^2 = 1,317/1,285^2 = 0,798.$$

$r^2 = 0,763^2 = 0,582$. Характеристика нелинейности в данном случае $\eta^2 - r^2 = 0,798 - 0,582 = 0,216$.

Если соединить значения $y(x)$, обозначенные на рис. 13.2 светлыми точками, отрезками, то будет видно, что полученная ломаная линия незначительно отклоняется от прямой.

13.2. Многомерный корреляционный анализ

Если имеется многомерная нормально распределенная совокупность с n признаками X_1, X_2, \dots, X_n , то взаимозависимость между ними описывается корреляционной матрицей, под которой понимают матрицу, составленную из парных коэффициентов корреляции:

$$Q_n = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \dots & \rho_{1k} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} \dots & \rho_{2k} & \dots & \rho_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{j1} & \rho_{j2} & \rho_{j3} \dots & \rho_{jk} & \dots & \rho_{jn} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} \dots & \rho_{nk} & \dots & 1 \end{pmatrix},$$

где ρ_{jk} – парные коэффициенты корреляции.

В случае многомерной корреляции нельзя ограничиться одной корреляционной матрицей, так как зависимости между признаками сложны. Для более детального анализа используются частные коэффициенты корреляций различных порядков, позволяющие оценивать связь между двумя признаками при фиксированном значении остальных.

Если исходная совокупность состоит из n признаков, то частный коэффициент корреляции l -го порядка отражает зависимость между двумя из них при фиксированных значениях l признаков из $(n - 2)$ оставшихся. Если имеется, например, система из трех признаков X_1, X_2 и X_3 , то можно определять частные коэффициенты корреляции только первого порядка, так как в данном случае нельзя фиксировать больше одного признака. Если фиксировать значение X_3 , то

$$\rho_{12,3} = (\rho_{12} - \rho_{13}\rho_{23}) / \sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}, \quad (13.5)$$

где $\rho_{12,3}$ – частный коэффициент корреляции между признаками X_1 и X_2 при фиксированном значении X_3 ;

ρ_{12}, ρ_{13} , и ρ_{23} – парные коэффициенты корреляции.

Аналогично определяются $\rho_{13,2}$ и $\rho_{23,1}$.

Расчет частных коэффициентов позволяет оценить взаимное влияние признаков. Если, например, корреляция между X_1 и X_2 основана только на

общем влиянии X_3 , то $\rho_{12,3} = 0$. Если имеются четыре признака, то можно зафиксировать значения одного или двух признаков. В последнем случае

$$\rho_{12,34} = \frac{\rho_{12,4} - \rho_{13,4}\rho_{23,4}}{\sqrt{(1 - \rho_{13,4}^2)(1 - \rho_{23,4}^2)}} = \frac{\rho_{12,3} - \rho_{14,3}\rho_{24,3}}{\sqrt{(1 - \rho_{14,3}^2)(1 - \rho_{24,3}^2)}}.$$

Частные коэффициенты корреляции вычисляются на основании оценок парных коэффициентов корреляции. Так же, как и для парных, проверяется значимость частных коэффициентов корреляции, но при этом число степеней свободы при исключении каждого признака уменьшается на единицу.

Если фиксируется значение одного признака, то вместо $(m - 2)$ следует подставить $(m - 3)$, а если фиксируется значение двух признаков, то $(m - 4)$.

Для оценки линейной связи одного из признаков со всеми остальными используется множественный, или совокупный, коэффициент корреляции. Для случая трех признаков коэффициент множественной корреляции оценивается по формуле

$$R_{123} = \sqrt{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23} / (1 - r_{23}^2)},$$

при этом X_2 и X_3 , зависимость признака X_1 от которых оценивает R_{123} , считаются независимыми.

Для случая четырех признаков

$$R_{1234} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13,2}^2)(1 - r_{14,23}^2)}.$$

Значимость коэффициента множественной корреляции определяется при помощи критерия Фишера. Наблюдаемое значение критерия

$$F_{набл} = \frac{R^2}{1 - R^2} \cdot \frac{m - l - 2}{l},$$

где m – число наблюдений; $l = n - 1$, n – число признаков.

Критическое значение критерия Фишера определяется по таблице соответствующих критических значений согласно принятым доверительной вероятности P и числу степеней свободы.

Принимается $m_1 = f_1 + 1$ и $m_2 = f_2 + 1$.

Пример. При изучении бесцентрового шлифования роликов методом на проход контролировались следующие показатели качества: погрешность формы X_1 погрешность размера X_3 и параметр шероховатости X_2 . Всего было отшлифовано 50 образцов.

Путем измерения шероховатости и обработки результатов измерения была получена следующая матрица парных коэффициентов корреляции:

$$q_3 = \begin{pmatrix} 1 & 0,85 & 0,62 \\ & 1 & 0,53 \\ & & 1 \end{pmatrix}.$$

Необходимо установить характер взаимовлияния признаков.

Решение. Вычисляем:

$$r_{12,3} = (0,85 - 0,62 \cdot 0,53) / \sqrt{(1 - 0,62^2)(1 - 0,53^2)} = 0,784;$$

$$r_{23,1} = (0,53 - 0,85 \cdot 0,62) / \sqrt{(1 - 0,85^2)(1 - 0,62^2)} = 0,007;$$

$$r_{31,2} = (0,62 - 0,85 \cdot 0,53) / \sqrt{(1 - 0,85^2)(1 - 0,53^2)} = 0,379.$$

Проверяем значимость частных коэффициентов корреляции:

$$t_{набл1} = 0,784 \sqrt{47 / (1 - 0,784^2)} = 8,658;$$

$$t_{набл2} = 0,007 \sqrt{47 / (1 - 0,007^2)} = 0,048;$$

$$t_{набл3} = 0,379 \sqrt{47 / (1 - 0,379^2)} = 0,808.$$

Согласно таблице критических значений Стьюдента при $P = 0,95$ $t_{кр} = 2$. Поскольку $t_{набл1} > t_{кр}$ и $t_{набл3} > t_{кр}$, то гипотеза о незначимости связи между X_1 и X_2 , а также X_1 и X_3 отвергается. Но $t_{набл2} < t_{кр}$, поэтому величины X_2 и X_3 являются независимыми. Следовательно, условия шлифования, определяющие изменение погрешностей формы, влияют одновременно и на уровни значения параметра шероховатости и погрешности размера. Причем эти условия сильнее влияют на шероховатость, чем на погрешность размера, так как $r_{12,3} > r_{31,2}$.

13.3. Ранговая корреляция Спирмена

Иногда объекты генеральной совокупности обладают качественными признаками, т.е. признаками, которые невозможно измерить точно. Однако бывает важно установить, есть ли между объектами статистическая связь, например, связан ли характер стружкозавивания при точении с качеством обработанной поверхности или с условиями труда станочника. Сила связи между качественными признаками объектов носит название *ранговой корреляции*. Для практических целей использование ранговой корреляции весьма полезно, а иногда является единственно возможным способом объективной оценки. Например, если установлена высокая ранговая корреляция между двумя качественными признаками изделий, то достаточно контролировать изделия только по одному из признаков, что удешевляет и ускоряет контроль.

Для проведения оценки этой связи необходимо, в первую очередь, расположить объекты в порядке убывания или возрастания качества и в соответствии с местом, на котором объект окажется, присвоить ему ранг. При ранжировании в порядке ухудшения качества на первом месте будет находиться объект наилучшего качества по сравнению с остальными; на втором месте окажется объект хуже первого, но лучше других, и т.д.

Положим, что выборка объема n содержит независимые объекты, которые обладают двумя качественными признаками A и B . Для оценки степени связи признаков находят коэффициенты ранговой корреляции, в частности коэффициент Спирмена.

Расположим сначала объекты выборки в порядке ухудшения качества по признаку A . Припишем объекту, стоящему на i -м месте, число – ранг x_i ; равный порядковому номеру объекта, т.е. ранг объекта, занимающего первое место, будет равен 1, объект, расположенный на втором месте, получит ранг 2 и т.д. В итоге получим последовательность рангов по признаку A : $x_1 = 1, x_2 = 2, \dots, x_n = n$.

Расположим теперь объекты в порядке убывания качества по признаку B и припишем каждому из них ранг y_i , однако индекс i при y будет равен по прежнему порядковому номеру объекта по признаку A . Например, запись $y_2 = 5$ означает, что по признаку A объект стоит на втором месте, а по признаку B – на пятом. Если выборка содержит объекты с одинаковым качеством, то каждому из них приписывается ранг, равный среднему арифметическому порядковых номеров объектов.

В итоге получим две последовательности рангов:

по признаку A x_1, x_2, \dots, x_n ;

по признаку B y_1, y_2, \dots, y_n .

Заметим, что в первой строке индекс i совпадает с порядковым номером объекта, а во второй – не совпадает. На практике чаще будет встречаться промежуточный случай, когда ухудшение качества по одному признаку влечет за собой для некоторых объектов ухудшение, а для других – улучшение качества. Задача состоит в том, чтобы оценить связь между признаками. Для ее решения рассмотрим ранги x_1, x_2, \dots, x_n как возможные значения случайной величины X , а y_1, y_2, \dots, y_n – как возможные значения случайной величины Y . Таким образом, о связи между качественными признаками A и B можно судить по связи между случайными величинами X и Y , для оценки которой используем коэффициент корреляции:

$$r_{\rho} = \frac{\sum u_i v_i}{n \sigma_u \sigma_v}.$$

Оценка коэффициента корреляции проводится путем расчета выборочного коэффициента корреляции Спирмена ρ_s :

$$\rho_s = 1 - \left[\frac{6 \sum d_i^2}{n^3 - n} \right],$$

где $d_i = x_i - y_i$.

Если выборочный коэффициент ранговой корреляции Спирмена равен единице, то между качественными признаками A и B имеется полная прямая зависимость. Если же выборочный коэффициент ранговой корреляции

Спирмена равен -1 , то между качественными признаками A и B имеется противоположная зависимость.

Обычно коэффициент ρ_s заключен между -1 и $+1$, причем чем ближе к нулю его абсолютная величина, тем зависимость между качественными признаками A и B меньше.

Пример. Предположим, экспертам предложили 10 стружек, полученных при точении одной и той же заготовки на разных режимах обработки. Было предложено расположить образцы в порядке уменьшения следующих признаков (т.е. ранжировать): удобство автоматизированной транспортировки (признак x_i) и предполагаемая стойкость режущей пластины (признак y_i). Были получены следующие результаты ранжирования:

x_i	1	2	3	4	5	6	7	8	9	10
y_i	6	4	8	1	2	5	10	3	7	9

Определить выборочный коэффициент ранговой корреляции Спирмена по данным ранга объектов выборки объемом $n = 10$.

Решение. Находим разности рангов $d_i = x_i - y_i : -5, -2, -5, 3, 3, 1, -3, 5, 2, 1$.

Вычислим сумму квадратов разностей рангов:

$$\sum d_i^2 = 25 + 4 + 25 + 9 + 9 + 1 + 9 + 25 + 4 + 1 = 112.$$

Найдем искомый коэффициент ранговой корреляции, учитывая, что $n = 10$:

$$\rho_s = 1 - \left[6 \sum d_i^2 / (n^3 - n) \right] = 1 - \left[6 \cdot 112 / (1000 - 10) \right] = 0,32.$$

Выдвигаем нулевую гипотезу H_0 : генеральный коэффициент ранговой корреляции ρ_s Спирмена равен нулю при конкурирующей гипотезе $H_1: \rho_s \neq 0$.

Вычисляем критическое значение:

$$T_{кр} = t_{кр}(\alpha; k) \sqrt{(1 - \rho_s^2) / (n - 2)}, \quad (13.6)$$

где n – объем выборки; ρ_s – выборочный коэффициент ранговой корреляции Спирмена; $t_{кр}(\alpha; k)$ – критическая точка двусторонней критической области,

которую находят по таблице критических точек распределения Стьюдента по уровню значимости α и числу степеней свободы $k=n-2$.

Если $|\rho_B| < T_{кр}$, то нет оснований отвергнуть нулевую гипотезу, т.е. ранговая корреляционная связь между качественными признаками незначима.

Если $|\rho_B| > T_{кр}$, то нулевую гипотезу отвергают. Между качественными признаками существует значимая ранговая корреляционная связь.

По формулам (13.5) и (13.6) для рассматриваемого случая при уровне значимости 0,05 и числе степеней свободы $k = 10 - 2 = 8$ находим $t_{кр}(\alpha; k) = 2,31$. Следовательно, $T_{кр} = 0,773$, и так как $|\rho_B| < T_{кр}$, то нет оснований отвергнуть нулевую гипотезу, согласно которой ранговая корреляционная связь между признаками незначима.

Контрольные вопросы

1. Что оценивают коэффициенты корреляции?
2. В каких пределах может изменяться коэффициент корреляции?
3. Как строится поле корреляции?
4. С помощью какого критерия проверяется значимость выборочного коэффициента корреляции?
5. Что выявляет множественный коэффициент корреляции?
6. В каких случаях применяют ранговую корреляцию?

СПИСОК ЛИТЕРАТУРЫ

1. Адлер Ю.П. Введение в планирование эксперимента. М.: Металлургия, 1969. 270 с.
2. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. М.: Наука, 1976. 280 с.
3. Красовский Г.И., Филаретов Г.Ф. Планирование эксперимента. Мн.: Изд-во БГУ, 1982. 302 с.
4. Боровиков В.П., Боровиков И.П. STATISTICA: Статистический анализ и обработка данных в среде Windows. М.: Филинь, 1997. 608 с.
5. Рогов В.А., Позняк Г.Г. Методика и практика технических экспериментов: учеб. пособие для студ. высш. учеб. заведений. М.: Издательский центр «Академия», 2005. 288 с.
6. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учеб. пособие. М.: ФОРУМ: ИНФРА, 2004. 464 с.
7. Чекмарев Н.А., Барвинок В.А., Шалавин В.В. Статистические методы управления качеством. М.: Машиностроение, 1999. 320 с.
8. Математическое планирование эксперимента в машиностроении / В.А. Бордаков, С.А. Бордаков, С.Ф. Демичев [и др.]. Самара: Самар. гос. аэрокосм. ун-т, 2002. 56 с.
9. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. Методы планирования эксперимента. М.: Мир, 1981. 520 с.
10. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. Методы обработки данных. М.: Мир, 1980. 610 с.
11. Налимов В.В., Голикова Т.И. Логическое основание планирования эксперимента. М.: Металлургия, 1980. 152 с.
12. Шенк Х. Теория инженерного эксперимента / пер. с нем. М.: Мир, 1972. 382 с.

13. Ящерицын П.И., Махаринский Е.И. Планирование эксперимента в машиностроении. Минск: Вышэйш. шк., 1985. 286 с.
14. Монтгомери Д.К. Планирование эксперимента и анализ данных. Л.: Судостроение, 1980. 384 с.
15. Анфилатов В.С., Емельянов А.А., Кукушкин А.А. Системный анализ в управлении. М.: Финансы и статистика, 2002. 242 с.
16. Асатурян В.И. Теория планирования эксперимента. М.: Радио и связь, 1983. 260 с.
17. Надежность и эффективность в технике. Т.6. Экспериментальная отработка и испытания. М.: Машиностроение, 1989. 264 с.
18. Налимов В.В. Теория эксперимента. М.: Наука, 1971. 164 с.
19. Ходасевич Г.Б. Обработка экспериментальных данных на ЭВМ. Часть 1. Обработка одномерных данных. СПб.: СПбГУТ, 2002. 216 с.
20. Ходасевич Г.Б. Обработка экспериментальных данных на ЭВМ. Часть 2. Обработка многомерных данных. СПб.: СПбГУТ, 2002. 224 с.
21. Таблицы планов эксперимента для факторных и полиномиальных моделей / под ред. В.В. Налимова. М.: Металлургия, 1982. 124 с.

ПРИЛОЖЕНИЯ

Приложение 1

Значение функции нормального распределения $\Phi(x)$ стандартного нормального закона $N(0,1)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt; \quad \Phi(x) \equiv 1 - \Phi(-x)$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,500000	1,00	0,841345	2,00	0,977250
0,05	0,519939	1,05	0,853141	2,05	0,979818
0,10	0,539828	1,10	0,864334	2,10	0,982136
0,15	0,559618	1,15	0,874928	2,15	0,984222
0,20	0,579260	1,20	0,884893	2,20	0,986097
0,25	0,589706	1,25	0,894350	2,25	0,987776
0,30	0,617911	1,30	0,903200	2,30	0,989276
0,35	0,636831	1,35	0,911492	2,35	0,990613
0,40	0,655422	1,40	0,919243	2,40	0,991802
0,45	0,673645	1,45	0,926471	2,45	0,992857
0,50	0,691463	1,50	0,933193	2,50	0,993790
0,55	0,708840	1,55	0,939429	2,55	0,994614
0,60	0,725747	1,60	0,945201	2,60	0,995339
0,65	0,742154	1,65	0,950528	2,65	0,995975
0,70	0,758036	1,70	0,955434	2,70	0,996533
0,75	0,773373	1,75	0,959941	2,75	0,997020
0,80	0,788145	1,80	0,964070	2,80	0,997445
0,85	0,802338	1,85	0,967843	2,85	0,997814
0,90	0,815940	1,90	0,971283	2,90	0,998134
0,95	0,828944	1,95	0,974412	2,95	0,998411
				3,00	0,998630

Квантили u_p нормального распределения $N(0,1)$

p	0,90	0,95	0,975	0,99	0,995	0,999	0,995
u_p	1,282	1,645	1,960	3,326	2,576	3,090	3,291

Критерий для исключения высказывающих вариант

Объем выборки, n	Уровень значимости					
	0,05			0,01		
	$\frac{x_n - x_{n-1}}{x_n - x_1}$	$\frac{x_n - x_{n-1}}{x_n - x_2}$	$\frac{x_n - x_{n-2}}{x_n - x_1}$	$\frac{x_n - x_{n-1}}{x_n - x_1}$	$\frac{x_n - x_{n-1}}{x_n - x_2}$	$\frac{x_n - x_{n-2}}{x_n - x_1}$
	$\frac{x_2 - x_1}{x_n - x_1}$	$\frac{x_2 - x_1}{x_n - x_2}$	$\frac{x_3 - x_1}{x_n - x_1}$	$\frac{x_2 - x_1}{x_n - x_1}$	$\frac{x_2 - x_1}{x_n - x_2}$	$\frac{x_3 - x_1}{x_n - x_1}$
3	0,941	1,000	1,000	0,998	1,000	1,000
4	0,765	0,955	0,967	0,889	0,991	0,992
5	0,642	0,807	0,845	0,780	0,916	0,929
6	0,560	0,689	0,736	0,698	0,805	0,836
7	0,507	0,660	0,661	0,637	0,740	0,778
8	0,468	0,554	0,607	0,590	0,683	0,710
9	0,437	0,512	0,565	0,555	0,635	0,667
10	0,412	0,477	0,531	0,527	0,597	0,632
11	0,392	0,450	0,504	0,502	0,566	0,603
12	0,376	0,428	0,481	0,482	0,541	0,579
15	0,338	0,381	0,430	0,438	0,486	0,522
20	0,300	0,334	0,372	0,391	0,430	0,464
24	0,281	0,309	0,347	0,367	0,400	0,434
30	0,260	0,283	0,322	0,341	0,369	0,402

Значения t -критерия Стьюдента при уровне значимости 0,10; 0,05; 0,01

$df(v)$	α		
	0,10	0,05	0,01
1	6,3138	12,706	63,657
2	2,9200	4,3027	9,9248
3	2,3534	3,1825	5,8409
4	2,1318	2,7764	4,6041
5	2,0150	2,5706	4,0321
6	1,9432	2,4469	3,7074
7	1,8946	2,3646	3,4995
8	1,8595	2,3060	3,3554
9	1,8331	2,2622	3,2498
10	1,8125	2,2281	3,1693
11	1,7959	2,2010	3,1058
12	1,7823	2,1788	3,0545
13	1,7709	2,1604	3,0123
14	1,7613	2,1448	2,9768
15	1,7530	2,1315	2,9467
16	1,7459	2,1199	2,9208
17	1,7396	2,1098	2,8982
18	1,7341	2,1009	2,8784
19	1,7291	2,0930	2,8609
20	1,7247	2,0860	2,8453
21	1,7207	2,0796	2,8314
22	1,7171	2,0739	2,8188
23	1,7139	2,0687	2,8073
24	1,7109	2,0639	2,7969
25	1,7081	2,0595	2,7874
26	1,7056	2,0555	2,7787
27	1,7033	2,0518	2,7707
28	1,7011	2,0484	2,7633
29	1,6991	2,0452	2,7564
30	1,6973	2,0423	2,7500
40	1,6839	2,0211	2,7045
60	1,6707	2,0003	2,6603
120	1,6577	1,9799	2,6174
∞	1,6449	1,9600	2,5758

Значения χ^2 -критерия Пирсона при уровне значимости 0,10; 0,05; 0,01

$df(v)$	0,10	0,05	0,01
1	2,71	3,84	6,63
2	4,61	5,99	9,21
3	6,25	7,81	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
11	17,28	19,68	24,72
12	18,55	21,03	26,22
13	19,81	22,36	27,69
14	21,06	23,68	29,14
15	22,31	25,00	30,58
16	23,54	26,30	32,00
17	24,77	27,59	33,41
18	25,99	28,87	34,81
19	27,20	30,14	36,19
20	28,41	31,41	37,57
21	29,62	32,67	38,93
22	30,81	33,92	40,29
23	32,01	34,17	41,64
24	33,20	36,42	42,98
25	34,38	37,65	44,31
26	35,56	38,89	45,64
27	36,74	40,11	46,96
28	37,92	41,34	48,28
29	39,09	42,56	49,59
30	40,26	43,77	50,89
40	51,80	55,76	63,69
50	63,17	67,50	76,15
60	74,40	79,08	88,38
70	85,53	90,53	100,42
80	96,58	101,88	112,33
90	107,56	113,14	124,12
100	118,50	124,34	135,81

Значения F -критерия Фишера при уровне значимости 0,05

df_1 (v_1)	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	30	∞
df_2 (v_2)																	
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	250	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,46	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74	8,71	8,69	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,74	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,57	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69	2,64	2,60	2,54	2,46	2,30
13	4,67	3,81	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	2,55	2,51	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53	2,48	2,44	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48	2,43	2,39	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42	2,37	2,33	2,28	2,20	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38	2,33	2,29	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31	2,26	2,21	2,15	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28	2,23	2,18	2,12	2,04	1,84

$df_1 (v_1)$	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	30	∞
$df_2 (v_2)$																	
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,20	2,15	2,09	2,00	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23	2,18	2,13	2,07	1,98	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20	2,14	2,10	2,04	1,96	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18	2,13	2,09	2,02	1,94	1,73
25	4,24	3,88	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,20	2,16	2,11	2,06	2,00	1,92	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,90	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13	2,08	2,03	1,97	1,88	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,87	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,05	2,00	1,94	1,85	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09	2,04	1,99	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00	1,95	1,90	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95	1,90	1,85	1,78	1,69	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,86	1,81	1,75	1,65	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85	1,79	1,75	1,68	1,57	1,28
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,79	1,75	1,69	1,64	1,57	1,46	1,00

Значения G-критерия Кохрена при уровне значимости 0,05

<i>n</i>	<i>f</i>													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010	0,7880	0,7341	0,6602	0,5813	0,5000
3	0,9669	0,8709	0,7977	0,7457	0,7071	0,6771	0,6530	0,6333	0,6167	0,6025	0,5466	0,4748	0,4031	0,3333
4	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017	0,4884	0,4366	0,3720	0,3093	0,2500
5	0,8412	0,6838	0,5981	0,5440	0,5063	0,4783	0,4564	0,4387	0,4241	0,4118	0,3645	0,3066	0,2513	0,2000
6	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682	0,3568	0,3135	0,2612	0,2119	0,1667
7	0,7271	0,5612	0,4800	0,4307	0,3974	0,3726	0,3535	0,3384	0,3259	0,3154	0,2756	0,2278	0,1833	0,1429
8	0,6798	0,5175	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926	0,2829	0,2462	0,2022	0,1616	0,1250
9	0,6385	0,4775	0,4027	0,3584	0,3286	0,3067	0,2901	0,2768	0,2659	0,2568	0,2226	0,1820	0,1446	0,1111
10	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,2666	0,2541	0,2439	0,2353	0,2032	0,1655	0,1308	0,1000
12	0,5410	0,3924	0,3264	0,2880	0,2624	0,2439	0,2299	0,2187	0,2098	0,2020	0,1737	0,1403	0,1100	0,0833
15	0,4709	0,3346	0,2758	0,2419	0,2195	0,2034	0,1911	0,1815	0,1736	0,1671	0,1429	0,1144	0,0889	0,0667
20	0,3894	0,2705	0,2205	0,1921	0,1735	0,1602	0,1501	0,1422	0,1357	0,1303	0,1108	0,879	0,0675	0,0500
24	0,3434	0,2354	0,1907	0,1656	0,1493	0,1374	0,1286	0,1216	0,1160	0,1113	0,0942	0,0743	0,0567	0,0417
30	0,2929	0,1980	0,1593	0,1377	0,1237	0,1137	0,1061	0,1002	0,0958	0,0921	0,0771	0,0604	0,0457	0,0333
40	0,2370	0,1576	0,1259	0,1082	0,0968	0,0887	0,0827	0,0780	0,0745	0,0713	0,0595	0,0462	0,0347	0,0250
60	0,1737	0,1131	0,0895	0,0765	0,0682	0,0623	0,0583	0,0552	0,0520	0,0497	0,0411	0,0316	0,0234	0,0167
120	0,0998	0,0632	0,0495	0,0419	0,0371	0,0337	0,0312	0,0292	0,0279	0,0266	0,0218	0,0165	0,0120	0,0083



Процентные точки F -распределения

Величины $t_{v_1, v_2, P}$ таковы, что

$$P = \frac{1}{B(v_1/2, v_2/2)} \int_0^{v_1 F_{v_1, v_2, p/v_2}} g^{v_1/2-1} (1+g)^{-(v_1+v_2)/2} dg.$$

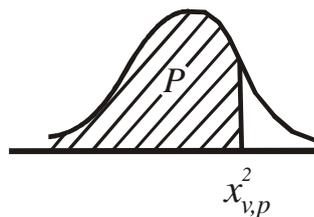
$F = 0,95.$

P v	1	2	3	4	5	6	7	8	9	10	12	15	20	20	30	40	60	120	∞
1	661,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,44	19,45	19,46	19,47	19,48	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,19	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,58	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,8	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Процентные точки χ^2 -распределения

Величины $\chi^2_{v,P}$ таковы, что

$$P = \frac{1}{2^{v/2}\Gamma(v/2)} \int_0^{\chi^2_{v,P}} y^{v/2-1} e^{-y/2} dy.$$



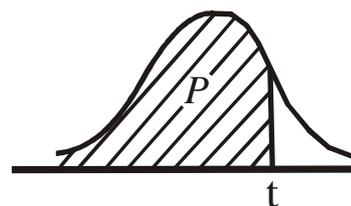
<i>P</i>	0,005	0,010	0,025	0,050	0,100	0,250	0,500
1	0,00004	0,00016	0,00098	0,00393	0,01579	0,1015	0,4549
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,5754	1,386
3	0,0717	0,1148	0,2158	0,3518	0,5844	1,213	2,366
4	0,2070	0,2971	0,4844	0,7107	1,064	1,923	3,357
5	0,4117	0,5543	0,8312	1,145	1,610	2,675	4,351
6	0,6757	0,8721	1,2373	1,635	2,204	3,455	5,348
7	0,9893	1,239	1,690	2,167	2,833	4,255	6,346
8	1,344	1,646	2,180	2,733	3,490	5,071	7,344
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342
11	2,603	3,053	3,816	4,575	5,578	7,584	10,34
12	3,074	3,571	4,404	5,226	6,304	8,438	11,34
13	3,565	4,107	5,009	5,892	7,041	9,299	12,34
14	4,075	4,60	5,629	6,571	7,790	10,17	13,34
15	4,601	5,229	6,262	7,261	8,547	11,04	14,34
16	5,142	5,812	6,908	7,962	9,312	11,91	15,34
17	5,697	6,408	7,564	8,672	10,09	12,79	16,34
18	6,265	7,015	8,231	9,390	10,86	13,68	17,34
19	6,844	7,633	8,907	10,12	11,65	14,56	18,34
20	7,434	8,260	9,591	10,85	12,44	15,45	19,34
21	8,034	8,897	10,28	11,59	13,24	16,34	20,34
22	8,643	9,542	10,98	12,34	14,04	17,24	21,34
23	9,260	10,20	11,69	13,09	14,85	18,14	22,34
24	9,886	10,86	12,40	13,85	15,66	19,04	23,34
25	10,52	11,52	13,12	14,61	16,47	19,94	24,34
26	11,16	12,20	13,84	15,38	17,29	20,84	25,34
27	11,81	12,88	14,57	16,15	18,11	21,75	26,34
28	12,46	13,56	15,31	16,93	18,94	22,66	27,34
29	13,12	14,26	16,05	17,71	19,77	23,57	28,34
30	13,79	14,95	16,79	18,49	20,60	24,48	29,34
40	20,71	22,16	24,43	26,51	29,05	33,66	39,34
50	27,99	29,71	32,36	34,76	37,69	42,94	49,33
60	35,53	37,48	40,48	43,19	46,46	52,29	59,33
70	43,28	45,44	48,76	51,74	55,33	61,70	69,33
80	51,17	53,54	57,15	60,39	64,28	71,14	79,33
90	59,20	61,75	65,65	69,13	73,29	80,62	89,33
100	67,33	70,06	74,22	77,93	82,36	90,13	99,33

<i>P</i>	0,750	0,900	0,950	0,975	0,990	0,995	0,999
<i>v</i>							
1	1,323	2,706	3,841	5,024	6,635	7,879	10,89
2	2,773	4,605	5,991	7,378	9,210	10,60	13,82
3	4,108	6,251	7,815	9,348	11,34	12,84	16,27
4	5,385	7,779	9,488	11,14	13,28	14,86	18,47
5	6,626	9,236	11,07	12,83	15,09	16,75	20,52
6	7,841	10,64	12,59	14,45	16,81	18,55	22,46
7	9,037	12,02	14,07	16,01	18,48	20,28	24,32
8	10,22	13,36	15,51	17,53	20,09	21,96	26,12
9	11,39	14,68	16,92	19,02	21,67	23,59	27,88
10	12,55	15,99	18,31	20,48	23,21	25,19	29,59
11	13,70	17,28	19,68	21,92	24,72	26,76	31,26
12	14,85	18,55	21,03	23,34	26,22	28,30	32,91
13	15,98	19,81	22,36	24,74	27,69	29,82	34,53
14	17,12	21,06	23,68	26,12	29,14	31,32	36,12
15	18,25	22,31	25,00	27,49	30,58	32,80	37,70
16	19,37	23,54	26,30	28,85	32,00	34,27	39,25
17	20,49	24,77	27,59	30,19	33,41	35,72	40,79
18	21,60	25,99	28,87	31,53	34,81	37,16	42,31
19	22,72	27,20	30,14	32,85	36,19	38,58	43,82
20	23,83	28,41	31,41	34,17	37,57	40,00	45,32
21	24,93	29,62	32,67	35,48	38,93	41,40	46,80
22	26,04	30,81	33,92	36,78	40,29	42,80	48,27
23	27,14	32,01	35,17	38,08	41,64	44,18	49,73
24	28,24	33,20	36,42	39,36	42,98	45,56	51,18
25	29,34	34,38	37,65	40,65	44,31	46,93	52,62
26	30,43	35,56	38,89	41,92	45,64	48,29	54,05
27	31,53	36,74	40,11	43,19	46,96	49,64	55,48
28	32,62	37,92	41,34	44,46	48,28	50,99	56,89
29	33,71	39,09	42,56	45,72	49,59	52,34	58,30
30	34,80	40,26	43,77	46,98	50,89	53,67	59,70
40	45,62	51,80	55,76	59,34	63,69	66,77	73,40
50	56,33	63,17	67,50	71,42	76,15	79,49	86,66
60	66,98	74,40	79,08	83,30	88,38	91,95	99,61
70	77,58	85,53	90,53	95,02	100,4	104,2	112,3
80	88,13	96,58	101,9	106,6	112,3	116,3	124,8
90	98,65	107,6	113,1	118,1	124,1	128,3	137,2
100	109,1	118,5	124,3	129,6	135,8	140,2	149,4

Процентные точки t -распределения Стьюдента

Величины $t_{v,P}$ таковы, что

$$P = \int_{-\infty}^{t_{v,P}} \frac{1}{\sqrt{v\pi}} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}} dt.$$



P	0,750	0,900	0,950	0,975	0,990	0,995	0,999	0,9995
v								
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31	636,62
2	0,816	1,886	2,920	4,303	6,965	9,925	22,326	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,213	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	2,055	3,930	4,4318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,32	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232	3,460
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160	3,373
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090	3,291